

THE CHARACTERIZATION OF THE *mier1* MLP-P1  
PROMOTER; ITS ACTIVITY AND IMPLICATIONS IN  
BREAST CANCER CELLS

JACLYN A. CLEMENTS







THE CHARACTERIZATION OF THE *mier1* MLP-P1 PROMOTER;  
ITS ACTIVITY AND IMPLICATIONS IN BREAST CANCER CELLS

By Jaclyn A. Clements

A thesis submitted to the  
School of Graduate Studies  
in partial fulfillment of the  
requirements for the degree of  
Masters of Science

Division of BioMedical Sciences

Faculty of Medicine

Memorial University

July 2010

St. John's

Newfoundland



## Abstract

Mesoderm induction early response 1 (*mier1*) is a gene encoding MIER1 $\alpha$ , a corepressor of ER $\alpha$  (Mercer *et al.*, unpublished data). As breast tumours become invasive, the subcellular localization of MIER1 $\alpha$  changes from the nucleus to the cytoplasm, thereby potentially abrogating or changing its function, an action of which may have a role in breast cancer progression (McCarthy *et al.*, 2008). This study analyzed the regulation of *mier1* in breast cancer cells by investigating the previously uncharacterized MLP-P1 promoter. The subsequent role of this promoter in the subcellular localization of MIER1 $\alpha$  was then studied in order to further elucidate possible mechanisms underlying MIER1 $\alpha$ 's documented change in subcellular localization. Transient transfection of a MLP-P1 promoter deletion series driving a luciferase reporter gene in multiple cell lines was used to characterize the minimum promoter and revealed that MLP-P1 is the predominately active *mier1* promoter. The degree to which its activity was increased over the second *mier1* promoter, MAEP-P2, was largest in ER+ breast cancer cell lines.

Transcriptional activation at MLP-P1 can lead to the incorporation of an alternate exon that generates transcripts harbouring an extra 74bp insert containing either a putative nuclear export sequence (NES) or transmembrane domain. Immunocytochemical analysis showed that MIER1 $\alpha$  isoforms resulting from transcripts harbouring this insert (denoted MIER1 3A  $\alpha$ ) localized to the cytoplasm, while MIER1 $\alpha$  isoforms lacking this extra amino acid sequence (denoted MIER1  $\alpha$ ) localized to the nucleus in ER+ MCF7 breast cancer cells. Leptomycin B (LMB), an inhibitor of the nuclear export system, decreased MIER1 3A  $\alpha$

cytoplasmic localization while simultaneously increasing its nuclear localization, thereby demonstrating that exon 3A encodes a functional NES. Interestingly, both MIER1 $\alpha$  isoforms were equally localized throughout the entire cell in ER- cell lines, a phenomenon which was not affected by LMB treatment. Thus, preferential activation of MLP-P1 in ER+ rather than in ER- breast cancer cells may now be considered a prime candidate mechanism driving the subcellular localization change of MIER1 $\alpha$  from the nucleus to the cytoplasm.



## Acknowledgements

“Science is not belief, but the will to find out”. ~*Anonymous*

“Somewhere, something incredible is waiting to be known” ~*Dr. Carl Sagan*  
(*American Astronomer, Writer, & Scientist, 1934-1996*)

“The most beautiful thing we can experience is the mysterious. It is the source of all true art and science”. ~ *Albert Einstein*

I would like to take this opportunity to extend my utmost appreciation and gratitude to my supervisor Dr. Laura Gillespie for everything she has done for me over the duration of my master's degree. From her unwavering and unyielding support and encouragement, to her expert guidance and teaching, to the act of just always being there whenever I needed assistance; I cannot thank her enough. I'd also like to thank Dr. Gary Paterno for also helping me to further understand the theories and principles behind certain areas of my project, as well as its methodology. Moreover, I'd like to extend my gratitude to my supervisory committee members Dr. Ken Kao and Dr. Robert Gendron for helping me to really focus on the essential questions and direction of my work and for their council throughout the duration of my project. Furthermore I'd also like to thank Corinne Mercer and Heather Fifield who both really helped me with many aspects of my research from technical troubleshooting to the mastering of various protocols and techniques. As well, I'd also like to send out a huge thanks the rest of the Terry Fox Crew (Amanda, Youlian, Zhjian, Phil, Mark, Andrew, and Keelia) for always being there and extending a hand whenever I needed it. And finally I would like to thank my family and all my friends whose unfaltering

confidence and belief in me never let my determination or persistence fade. Thanks again so much, to all of you, for helping to make this project and thesis a reality.



## Table of Contents

<b>Abstract.....</b>	<b>ii</b>
<b>Acknowledgements.....</b>	<b>iv</b>
<b>Table of Contents.....</b>	<b>vi</b>
<b>List of Figures.....</b>	<b>xiii</b>
<b>List of Tables.....</b>	<b>xvi</b>
<b>List of Abbreviations.....</b>	<b>xviii</b>
<b>Chapter 1: Introduction.....</b>	<b>1</b>
<b>1.1 The Nature of Cancer &amp; Its Development.....</b>	<b>1</b>
1.1.1 The Complexity of Cancer.....	1
1.1.2 Cancer Today.....	2
1.1.3 Cancer at the Cellular & Molecular Level.....	3
<b>1.2 Gene Expression.....</b>	<b>8</b>
1.2.1 Transcription.....	10
1.2.1.1 Promoter Regions.....	11
1.2.2 RNA Splicing.....	13
<b>1.3 Regulation of Gene Expression.....</b>	<b>14</b>
1.3.1 Regulation of Transcription.....	14
1.3.1.1 Promoter Regions & Regulation of Transcription..	16
1.3.1.2 Alternate Promoter Usage.....	16

<b>1.4 The Consequences of the Dysregulation of Gene Expression: A Focus on Carcinogenesis.....</b>	<b>21</b>
1.4.1 Abnormal Gene Expression & Cancer Development.....	21
1.4.2 Abnormal Promoter Usage & Cancer Development.....	21
1.4.3 Aberrant Transcription Factor & Coregulator Signalling in Cancer Development.....	22
<b>1.5 Mesoderm Induction Early Response 1 (<i>mier1</i>).....</b>	<b>26</b>
1.5.1 <i>mier1</i> Gene, Transcripts, & Protein Isoform Structure.....	27
1.5.2 <i>mier1</i> promoters.....	32
1.5.2.1 MAEP-P2.....	32
1.5.2.2 MLP-P1.....	39
1.5.3 MIER1 Function.....	42
1.5.3.1 MIER1 as a Transcriptional Activator.....	42
1.5.3.2 MIER1 as a Transcriptional Repressor.....	44
1.5.3.2.1 Transcription Factor Tethering via the SANT Domain.....	44
1.5.3.2.2 Recruitment of HDAC1 via the ELM2 Domain.....	46
1.5.3.2.3 Estrogen Receptor (ER) Corepressor function of MIER1 $\alpha$ and implications for the LXXLL and the SANT Domain.....	47
1.5.3.2.4 Negative Regulation of the Histone Acetyltransferase (HAT) activity of CREB binding protein via MIER1 $\beta$ 's ELM2 Domain.....	49
1.5.4 MIER1 Expression & Subcellular Localization.....	50
1.5.5 Breast Cancer & MIER1.....	51
1.5.6 Regulation of <i>mier1</i> .....	53



1.6 The Aims and Principle Objectives of this Study.....	54
<b>Chapter 2: Materials &amp; Methods.....</b>	<b>57</b>
2.1 Cell Culture.....	57
2.1.1 Cell lines & Cell Maintenance.....	57
2.2 Vectors & Constructs.....	61
2.2.1 Experimental Control Plasmid Constructs.....	61
2.2.2 pCS3+MT <i>mier1</i> $\alpha$ & pCS3+MT <i>mier1</i> $\alpha$ exon3A.....	62
2.2.3 <i>mier1</i> Luciferase Reporter Gene Deletion Constructs.....	62
2.2.3.1 MLP-P1 Luciferase Reporter Gene Deletion Constructs.....	66
2.2.3.2 MAEP-P2 Luciferase Reporter Gene Deletion Constructs.....	69
2.3 Plasmid Construct Preparation, Purification, Quantification, & Storage.....	72
2.4 Reporter Gene Deletion Construct Design, Cloning, & Construction.....	73
2.4.1 <i>in silico</i> Promoter Analysis & Primer Design.....	75
2.4.2 Polymerase Chain Reaction.....	78
2.4.3 TOPO Cloning of <i>mier1</i> MLP-P1 Promoter Specific Insertion Sequences into the pCR2.1 Vector.....	81
2.4.4 Colony Screening, Mini-Preparation & Purification, DNA Quantification, and Sequencing of MLP-P1 (-91) pCR2.1 and MLP-P1 (-44) pCR2.1.....	82
2.4.5 Restriction Digest, Gel excision, Ligation, and Directional Cloning of MLP-P1 (-91) pCR2.1 & MLP-P1 (-44) pCR2.1 inserts into the pGL3 empty vector.....	86
2.4.6 Colony Screening, Mini-Preparation & Purification, DNA Quantification, and Sequencing of MLP-P1 (-91) pGL3 and MLP-P1(-44) pGL3.....	88

<b>2.5 Transient Transfection.....</b>	<b>88</b>
2.5.1 HEK 293, MDA MB 231, T47D, and Hs578T Cell Transfection.....	89
2.5.2 MCF7 Cell Transient Transfection.....	92
<b>2.6 Cell Lysis &amp; Protein Extraction.....</b>	<b>93</b>
<b>2.7 Reporter Assays.....</b>	<b>94</b>
2.7.1 Beta-Galactosidase ( $\beta$ -gal) Assay.....	94
2.7.2 Luciferase Assay.....	97
2.7.3 Bio-Rad Protein Assay.....	98
<b>2.8 Immunocytochemistry &amp; Transient Transfection.....</b>	<b>98</b>
2.8.1 Leptomycin B Treatment.....	106
<b>2.9 Statistical &amp; <i>in silico</i> Analyses.....</b>	<b>106</b>
 <b>Chapter 3: Results.....</b>	 <b>108</b>
<b>3.1 Characterization of MLP-P1 Promoter Activity in Breast Cancer Cells.....</b>	<b>108</b>
3.1.1 The Minimal Promoter of MLP-P1 in MCF7 Breast Cancer Cells.....	113
3.1.2 Maximal Promoter Activity at MLP-P1 in MCF7 Breast Cancer Cells.....	113
<b>3.2 Comparison of MLP-P1 Promoter Activity Patterns Across Multiple Breast Cancer Cell Lines varying in ER Status and the Non-Cancerous Cell Line HEK 293.....</b>	<b>116</b>
3.2.1 MLP-P1 Promoter Activity Patterns in HEK 293 Cells.....	116
3.2.2 MLP-P1 Promoter Activity Patterns in Multiple Breast Cancer Cell Lines varying in ER Status.....	120
3.2.3 MLP-P1 Promoter Activity in HEK 293 Cells compared to Multiple Breast Cancer Cell Lines varying in ER Status.....	124



<b>3.3 <i>In Silico</i> Analysis of the MLP-P1 Promoter Proximal Region.....</b>	<b>127</b>
3.3.1 Putative MLP-P1 Transcription Factor Binding Sites (TFBSs).....	127
3.3.2 Analysis of the Percent Homology of the MLP-P1 Promoter Proximal Sequence Conservation Across Various Species.....	133
<b>3.4 Comparison of MLP-P1 vs MAEP-P2 Promoter Activity in Breast Cancer Cells.....</b>	<b>137</b>
3.4.1 MAEP-P2 Promoter Activity Pattern in Breast Cancer Cells.....	138
3.4.2 MLP-P1 vs MAEP-P2 Promoter Activity in Breast Cancer Cells.....	145
<b>3.5 Effect of Incorporation of Exon 3A on MIER1<math>\alpha</math> Subcellular Localization.....</b>	<b>149</b>
3.5.1 Subcellular Localization of <i>mier1</i> Isoforms derived from Transcripts either containing or lacking the 74bp Insert encoded by Exon 3A in MCF7 Breast Carcinoma Cells.....	149
3.5.2 Subcellular Localization of <i>mier1</i> Isoforms derived from Transcripts either containing or lacking the 74bp Insert encoded by Exon 3A following Leptomycin B Treatment in MCF7 Cells.....	154
3.5.3 Subcellular Localization of <i>mier1</i> Isoforms derived from Transcripts either containing or lacking the 74bp Insert encoded by Exon 3A following Leptomycin B Treatment in HEK 293 Cells and MDA MB 231 Cells.....	157
<b>Chapter 4: Discussion.....</b>	<b>161</b>
<b>4.1 <i>mier1</i>'s MLP-P1 Promoter Region and Promoter Activity in Breast Cancer Cells.....</b>	<b>163</b>
4.1.1 The Minimal Promoter of MLP-P1.....	163
4.1.2 Maximum Activity Promoter Region of MLP-P1.....	165

4.2 Effect of ER Status on MLP-P1 Promoter Activity in Breast Cancer Cells.....	167
4.3 MLP-P1 Promoter activity in breast cancer cells vs. the non-cancerous HEK 293 cell line.....	173
4.4 Putative Transcription Factors involved in the Regulation of <i>mier1</i> at the MLP-P1 Promoter.....	177
4.5 MLP-P1 Promoter Proximal Sequence Conservation Across Various Species.....	188
4.6 MLP-P1 vs. MAEP-P2 Promoter Activity.....	189
4.7 Implications for Preferential Transcriptional Activation at MLP-P1 and the Effect of MLP-P1 Specific Exon 3A Inclusion on the Subcellular Localization of MIER1 $\alpha$ .....	193
4.8 Final Summary of Conclusions & Implications.....	202
References.....	206
Appendices.....	226
Appendix 1: <i>mier1</i> MLP-P1 Promoter Region showing location of 5' starting points for insertion sequences that were cloned into MLP-P1 Luciferase Reporter Gene Deletion Constructs.....	226
Appendix 2: <i>mier1</i> MLP-P1 Region Intron Sequence.....	228
Appendix 3: <i>mier1</i> MAEP-P2 Promoter Region & Location of Primers used to engineer Luciferase Reporter Gene Deletion Constructs.....	230
Appendix 4: Preliminary Results of the Characterization of the MLP-P1 Promoter Proximal Region.....	232
Appendix 5: MAEP-P2 Activity in HEK 293 Cells Recapitulates Previously Established Results Characterizing MAEP-P2 Promoter Activity in Alternate Cell Lines.....	235
Appendix 6: Comparison of the MAEP-P2 Promoter Activity Across Multiple Breast Cancer Cell Lines varying in ER Status and the HEK 293 Non-Cancerous Cell Line including the MAEP-P2 (+28) pGL3 Construct Results.....	236



<b>Appendix 7: Human MLP-P1 Promoter Sequence Aligned to Various Species.....</b>	<b>237</b>
---	------------

## List of Figures

Figure 1.1:	The six hallmarks of cancer.....	5
Figure 1.2:	The hallmarks of cancer development, including the stress phenotypes of cancer.....	6
Figure 1.3:	The Central Dogma of molecular biology.....	9
Figure 1.4:	Alternate RNA splicing and promoter usage.....	19
Figure 1.5:	<i>mier1</i> Gene Structure.....	28
Figure 1.6:	<i>mier1</i> Transcript & Protein Isoform Combinations.....	31
Figure 1.7:	MAEP-P2 promoter structure.....	34
Figure 1.8:	Characterization of MAEP-P2 promoter activity.....	36
Figure 1.9:	MAEP-P2 promoter CpG island sequence & putative transcription factor binding sites.....	38
Figure 1.10:	MLP-P1 promoter structure.....	40
Figure 1.11:	MLP-P1 proximal promoter sequence.....	41
Figure 1.12:	Structure of <i>mier1</i> protein isoforms including protein functional domains.....	43
Figure 2.1:	Schematic of <i>mier1</i> MLP-P1 promoter region with reference to the sequences incorporated into the MLP-P1 luciferase reporter gene deletion constructs.....	67
Figure 2.2:	DNA sequence of MLP-P1 promoter proximal region.....	68
Figure 2.3:	Schematic of <i>mier1</i> MAEP-P2 promoter region with reference to the sequences incorporated into the MAEP-P2 luciferase reporter gene deletion constructs.....	70
Figure 2.4:	DNA sequence of MAEP-P2 promoter proximal region.....	71
Figure 2.5:	Flow-chart for reporter gene deletion construct design, cloning, & construction.....	74



Figure 3.1:	Schematic of luciferase reporter gene deletion constructs used for the characterization of the minimum and maximum promoter activity regions of MLP-P1 in MCF7 breast cancer cells .....	112
Figure 3.2:	Characterization of the MLP-P1 promoter region of <i>mier1</i> in MCF7 cells.....	115
Figure 3.3:	Characterization of the MLP-P1 proximal promoter region of <i>mier1</i> in HEK 293 cells.....	119
Figure 3.4:	MLP-P1 promoter activity across multiple breast cancer cell lines varying in ER status.....	122
Figure 3.5:	MLP-P1 promoter activity in MDA MB 231 cells.....	123
Figure 3.6:	Comparison of the MLP-P1 promoter activity across multiple breast cancer cell lines varying in ER status and the non-cancerous HEK 293 cell line.....	126
Figure 3.7:	The sequence of the 94bp stretch of nucleotides residing between the 5' starting positions of the MLP-P1 promoter region inserts cloned into MLP-P1 (-185) pGL3 and MLP-P1 (-91) pGL3.....	128
Figure 3.8:	Putative human transcription factor binding sites in the <i>mier1</i> MLP-P1 promoter proximal region.....	132
Figure 3.9:	Graphical depiction of sequence homology of <i>mier1</i> 's MLP-P1 promoter region across six different species compared to the human sequence.....	136
Figure 3.10:	Schematic of luciferase reporter gene deletion constructs used for the initial characterization of MAEP-P2 promoter activity in MCF7 cells.....	140
Figure 3.11:	Characterization of the MAEP-P2 promoter region of <i>mier1</i> in MCF7 breast cancer cell line.....	141
Figure 3.12:	Analysis of the MAEP-P2 promoter activity across multiple breast cancer cell lines varying in ER status and the HEK 293 non-cancerous cell line.....	144
Figure 3.13:	Comparison of MLP-P1 vs MAEP-P2 promoter activity across multiple breast cancer cell lines varying in ER status and the non-cancerous HEK 293 cell line.....	147

Figure 3.14:	MLP-P1(-185) pGL3 vs. MAEP-P2 pGL3 luciferase activity comparison across various breast cancer cell lines and HEK 293 cells .....	148
Figure 3.15:	Subcellular localization data of <i>mier1</i> isoforms derived from transcripts either containing or lacking the 74bp insert encoded by exon 3A in MCF7 breast carcinoma cells.....	151
Figure 3.16:	Subcellular localization pictures of <i>mier1</i> isoforms containing or lacking the 74bp insert encoded by exon 3A in MCF7 breast carcinoma cells.....	153
Figure 3.17:	Effect of Leptomycin B treatment on <i>mier1</i> isoform subcellular localization in MCF7 cells.....	156
Figure 3.18:	Effect of Leptomycin B treatment on <i>mier1</i> isoform subcellular localization in HEK 293 cells and MDA MB 231 cells.....	160
Figure A1:	Preliminary results of the characterization of <i>mier1</i> MLP-P1 promoter proximal region in HEK 293 cells.....	233
Figure A2:	Characterization of the MAEP-P2 proximal promoter region of <i>mier1</i> in HEK 293 cells.....	235



## List of Tables

Table 1.1:	Hallmark ascriptions of oncogenic transcription factors.....	24
Table 2.1:	Cell lines.....	57
Table 2.2:	Cell line optimal confluency & subcultivation ratios.....	60
Table 2.3:	Experimental control plasmids.....	61
Table 2.4:	Primers used for cloning & engineering of MLP-P1 reporter gene deletion constructs.....	64
Table 2.5:	Primers used for cloning & engineering of MAEP-P2 reporter gene deletion constructs.....	65
Table 2.6:	Primers designed for the construction of MLP-P1 (-91) pGL3 and MLP-P1 (-44) pGL3 luciferase reporter gene deletion constructs.....	77
Table 2.7:	Platinum <sup>®</sup> <i>Taq</i> DNA Polymerase High Fidelity master mix reagents.....	79
Table 2.8:	Thermocycler parameters for PCR using Platinum <sup>®</sup> <i>Taq</i> DNA Polymerase High Fidelity.....	80
Table 2.9:	<i>Taq</i> DNA Polymerase master mix reagents for colony screening.....	84
Table 2.10:	Thermocycler parameters for PCR colony screening using <i>Taq</i> DNA Polymerase.....	85
Table 2.11:	Cell line seeding densities for 6-well plates.....	90
Table 2.12:	Temperatures & sample aliquot amounts for $\beta$ -Gal assay.....	96
Table 2.13:	Cell line seeding densities for 8-well chamber slides.....	105
Table 3.1:	Putative human transcription factor binding sites in the MLP-P1 promoter proximal region residing between -185 and -91bp from the MLP-P1 putative TSS.....	129
Table 3.2:	Putative human transcription factor binding sites in the MLP-P1 proximal promoter region residing between -91 to -44bp from the MLP-P1 Putative TSS.....	130

Table 3.3:	Percent conservation of MLP-P1 maximal and minimal promoter regions across six species compared to human sequences.....	134
------------	---	-----



## List of Abbreviations

$\alpha$	alpha
$\beta$	beta
$\beta$ -gal	beta-galactosidase
$\beta$ -M-ETOH	beta-mercaptoethanol
aa	amino acid
amp	ampicillin
ATCC	American Type Culture Collection
BLAST	Basic Local Alignment Search Tool
bp	base pair
BRCA1	breast cancer 1, early onset
BSA	bovine serum albumin
CAGE	genome cap analysis of gene expression
CAT	chloramphenicol acetyltransferase
CBP	CREB binding protein
cDNA	complementary DNA
C/EBP $\beta$	CCAAT/ enhancer binding protein beta
ChIP	chromatin immunoprecipitation
CMV	cytomegalovirus
co-IP	co-immunoprecipitation
coREST	REST co-repressor
CREB	cAMP response element binding protein
DAB	3, 3'- diaminobenzidine
DCIS	ductal carcinoma <i>in situ</i>
DMEM	Dulbecco's Modified Eagle Medium
DNA	deoxyribonucleic acid
EDTA	ethylenediamine tetraacetic acid
<i>E. coli</i>	<i>Escherichia coli</i>
ELM2	EGL-27 and MTA1 homology 2
ER	estrogen receptor
FAS	fatty acid synthase
FBS	fetal bovine serum
FCS	fetal calf serum
GST	glutathione s-transferase
H3K-9	histone H3 lysine 9
HAT	histone acetyltransferase
HDAC	histone deacetylase
HDAC1	histone deacetylase 1
HEK	human embryonic kidney
ICC	immunocytochemistry
IDC	invasive ductal carcinoma
ILC	invasive lobular carcinoma
IP	immunoprecipitation
KLF	krüppel-like factors
kb	kilobase pair

LB	luria broth
LMB	leptomycin B
MAPK	mitogen activated protein kinase
<i>mier1</i>	mesoderm induction early response 1 (DNA/RNA)
MIER1	mesoderm induction early response 1 (protein)
mRNA	messenger ribonucleic acid
MTA1	metastasis associated antigen 1
Na-pyruvate	sodium pyruvate
NCBI	National Center for Biotechnology Information
NCoR	nuclear receptor corepressor
NES	nuclear export sequence
NLS	nuclear localisation signal
PAS	polyadenylation site
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
Pen/Strep	penicillin/ streptomycin
PI3K	phosphatidylinositol-3-kinase
PIC	pre-initiation complex
PKI	protein kinase A inhibitor
pre-mRNA	pre-messenger ribonucleic acid
PRF	phenol red free
pRSV- $\beta$ -gal	plasmid rous sarcoma virus beta galactosidase
PTEN	phosphatase and tensin homolog
Q-Bit	Quant-iT <sup>TM</sup> dsDNA BR Assay Kit
qPCR	quantitative PCR
RACE	Rapid Amplification of cDNA Ends
RB	retinoblastoma
REST	RE1 silencing transcription factor
REV-T	reticuloendotheliosis retrovirus
RLU	relative luciferase units
RNA	ribonucleic acid
RSV	rous sarcoma virus
RTK	receptor tyrosine kinase
RT-PCR	reverse transcription PCR
SAM-HRP	sheep-anti mouse horseradish peroxidase
SANT	<u>S</u> W13, <u>A</u> DA2, <u>N</u> -CoR, and <u>T</u> FIIB
SFM	serum free media
siRNA	small interference RNA
SMRT	Silencing Mediator for Retinoid and Thyroid Hormone Receptors
Sp1	specificity protein 1
Src-3	steroid receptor co-activator 3
TBP	TATA binding protein
Tet	tetracycline
TFBS	transcription factor binding site
TFF1	trefoil factor 1
TFIID	transcription factor IID
TSS	transcription start site

tk	thymidine kinase
USF	upstream stimulatory factor
UTR	untranslated region
UV	ultraviolet
<i>xmier1</i>	<i>Xenopus</i> mesoderm induction early response 1(DNA/RNA)
XMIR1	<i>Xenopus</i> mesoderm induction early response 1 (Protein)



## **Chapter 1- Introduction**

### **1.1 The Nature of Cancer & Its Development**

#### **1.1.1 The Complexity of Cancer**

The dawn of the new millennium has brought with it an explosion of scientific discovery. This rapid progress has not only been observed in the field of medicine, with its innovative therapeutic breakthroughs, but also with regards to the understanding of the molecular, cellular, and physiological intricacies of the human body as a whole. From the sequencing of the entire human genome and advancements in stem cell and developmental research, to the technological feats of improved bioinformatics and *in silico* analysis algorithms, it is clear that our knowledge is undoubtedly expanding. As these scientific hallmarks are reached, the nature of our biological framework becomes more characterized; however, we are further faced with the undeniable truth of the complexity of our existence. A complexity that extends, as it is now, beyond our complete comprehension, as we are just starting to uncover the plethora of signalling networks and interconnecting factors, both at the molecular & environmental level, that are involved in ensuring the functional homeostasis of our bodies.

Currently, there exists no other disease that clearly exemplifies this complexity as cancer. Even though the past two decades have witnessed tremendous advances in our understanding of the pathogenesis of this disease, scientists are still continually unfolding new intricacies to the pro-cancer contributing factors, triggers, and the interplay between the neoplastic milieu and its surrounding environment (Luo *et al.*, 2009). Each cancer is a

complex network of dynamically evolving spatial-temporal molecular interactions (Mees *et al.*, 2009). Therefore, not only are there a multitude of factors and acquired properties that converge to generate the manifested phenotypes of cancer cells, but each type of cancer and even similar cancer-type cases from patient to patient can vary in patterns of origin, and can contain distinct key elements driving their progression. For example, many recent large-scale sequencing studies of multiple cancers have demonstrated that the collection of somatic mutations present in various cancer types differ, with each respective tumour harbouring a complex combination of low-frequency mutations thought to drive the cancer phenotype (Cancer Genome Atlas Research Network, 2008; Ding *et al.*, 2008; Jones *et al.*, 2008; Parsons *et al.*, 2008; Sjoblom *et al.*, 2006; Wood *et al.*, 2007).

### **1.1.2 Cancer Today**

An estimated 171,000 new cases of cancer and 75,300 deaths from cancer will have occurred in Canada in 2009, which represents approximately 470 Canadians diagnosed each day last year with some form of cancer (Canadian Cancer Statistics, 2009). Three types of cancer account for the majority of new cases in each sex: prostate, lung, and colorectal in males, and breast, lung, and colorectal in females. Based on current incidence rates, 40% of Canadian women and 45% of men will develop cancer during their lifetimes and 1 out of every 4 Canadians will die from it (Canadian Cancer Statistics, 2009).



To retaliate against such grim statistics, and the accruing resistance and shortcomings to current cancer therapies, research efforts are actively searching for innovative means to comprehend the molecular mechanisms of this disease and transfer this knowledge into novel therapies (Baudot *et al.*, 2009). More particularly, in order to address the heterogeneous nature and infrastructure of cancer between patients, research and clinical practice have been progressively skewing towards the concept of a more individual and personalized approach to therapy (Overdevest *et al.*, 2009). Personalized medicine focuses on and takes into account the individual's own genetic composition for the prevention, detection, prognosis and treatment of disease (Overdevest *et al.*, 2009; Phan *et al.*, 2009; Jain *et al.*, 2002). The eventual use of the personalized medicine approach will offer many benefits that hope to overcome the weaknesses and deficiencies facing current therapy (Refer to review Overdevest *et al.*, 2009; Ginsburg *et al.*, 2001). Today, as our ability to reach personalized medicine and cancer management comes closer and closer, so continues the fundamental importance of the continual investigation into cancer progression and its manifestations at the basic molecular level (Phan *et al.*, 2009; Allison *et al.*, 2008).

### **1.1.3 Cancer at the Cellular & Molecular Level**

Cancer is a disease involving dynamic alterations in the genome whereby distinct properties are acquired through a multistep, mutagenic process that allow cells to escape essential cellular regulatory and control checkpoints (Hanahan & Weinberg, 2000; Luo *et*



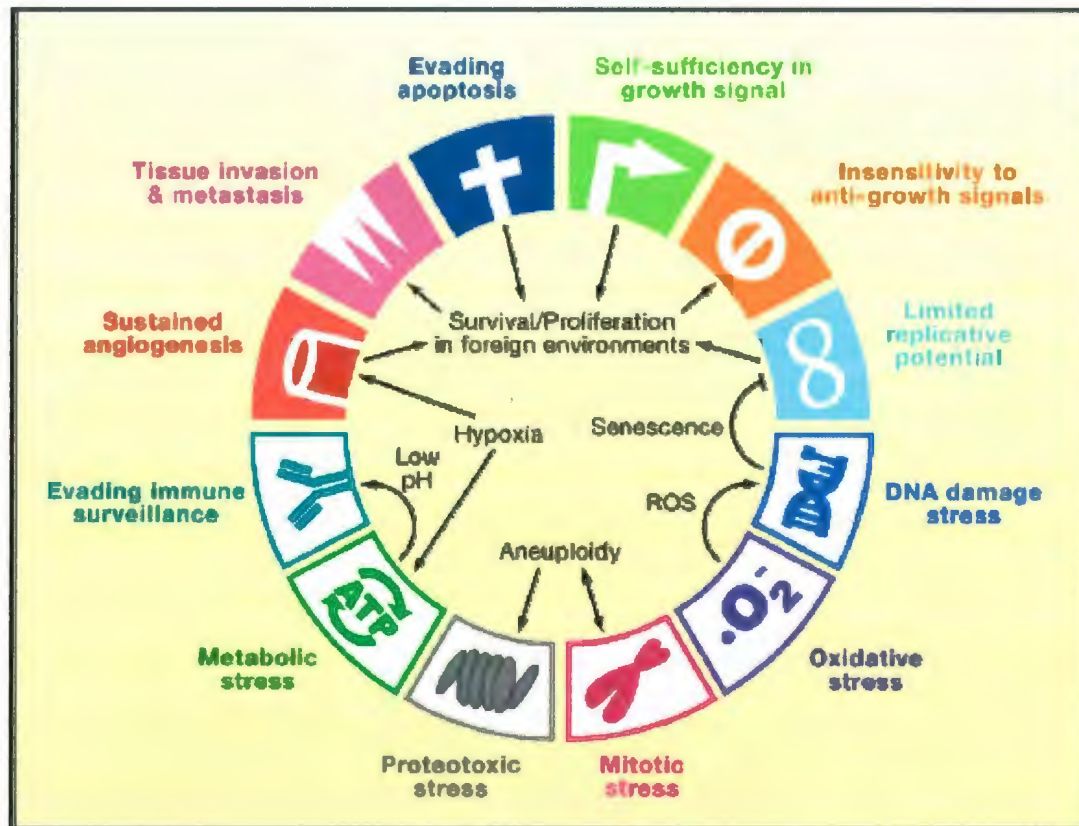
*al.*, 2009). The result of this subsequent liberation is a selfish, autonomously proliferating immortal cell that is completely noncompliant to the normal constraints governing cell growth & metabolism (Abelev & Erais, 2008). In 2000, Hanahan & Weinberg established a list of the six hallmarks of cancer that explain how normal human cells can transform into malignant cancers (Hanahan & Weinberg, 2000). Figure 1.1 depicts this list, which Hanahan & Weinberg, 2000 claim outline the essential cellular physiological alterations shared among all tumours that promote survival and proliferation. Recently, two additional groups have added further hallmarks representing the stress phenotypes of cancer (Fig 1.2) (Luo *et al.*, 2009; Kroemer & Pouyssegur, 2008). Together with the originally proposed hallmarks, these alterations collectively describe how cells can enter into the essential steps of the tumorigenic process, such as: gaining nutrients and oxygen, evading immune detection, attracting new blood vessels, and gaining support from surrounding stromal cells (Baudot *et al.*, 2009; Hanahan & Weinberg, 2000; Luo *et al.*, 2009).



**Figure 1.1: The six hallmarks of cancer**

This figure, from Hanahan & Weinberg, 2000<sup>1</sup> represents the six hallmarks of cancer : 1. Self-sufficiency in growth signals, 2. Insensitivity to anti-growth signals, 3. Evading apoptosis, 4. Sustained angiogenesis, 5. Limitless replicative potential, 6. Tissue invasion & metastasis.

<sup>1</sup> License agreement between JC and Elsevier provided by Copyright Clearance Center. License Number: 2391090086079. License date: March 16, 2010



**Figure 1.2: The hallmarks of cancer development, including the stress phenotypes of cancer**

This figure, from Luo *et al*, 2009<sup>2</sup>, depicts the aforementioned hallmarks of cancer development along with the additional stress phenotypes of cancer, which are: 1. Evading immune surveillance, 2. Metabolic stress, 3. Proteotoxic stress, 4. Mitotic stress, 5. Oxidative stress. Luo *et al*, 2009 assert that the functional interplay among these hallmarks promote the tumorigenic state and lead to the development of the originally proposed hallmarks, as illustrated by the demarked arrows. For example, the increased utilization of glycolysis that is often observed in tumor cells allows these cells to adapt to hypoxia and acidify the surrounding microenvironment, which further enables the evasion of the immune surveillance. Increased hypoxia furthermore supports sustained angiogenesis. Another interplay specifically depicted by this figure illustrates that increased mitotic stress can cause aneuploidy which can subsequently promote proteotoxic stress. Furthermore, elevated levels of reactive oxygen species results in increased levels of DNA damage, thereby further propagating the effects of tumor development (Luo *et al*, 2009).

<sup>2</sup> License agreement between JC and Elsevier provided by Copyright Clearance Center. License Number 2391090538555. License date: March 16, 2010



The progression of the aforementioned hallmarks in cancer is either directly or indirectly due to an accumulation of changes in the genome at the molecular level of a prospective cancer cell's deoxyribonucleic acid (DNA) sequence (Hanahan & Weinberg, 2000). These mutations and/ or structural changes are highly variable and can range from single point mutations in the DNA sequence, deletions and/ or insertions of specific nucleotides, to complex chromosomal rearrangements, translocations, fusions, and extensive changes in chromosome number (Luo *et al.*, 2009; Mees *et al.*, 2009). Carcinogenesis is initiated when such genomic modifications result in either a) improper activation and/ or repression or b) abnormal gain-of and/ or loss-of function of key genes involved in safeguarding and controlling important cellular processes (such as: cell growth and proliferation, cell survival, migration, apoptosis, polarity, differentiation during embryogenesis, and cellular homeostasis) (Luo *et al.*, 2009; Hahn & Weinberg, 2002, Mees *et al.*, 2009).

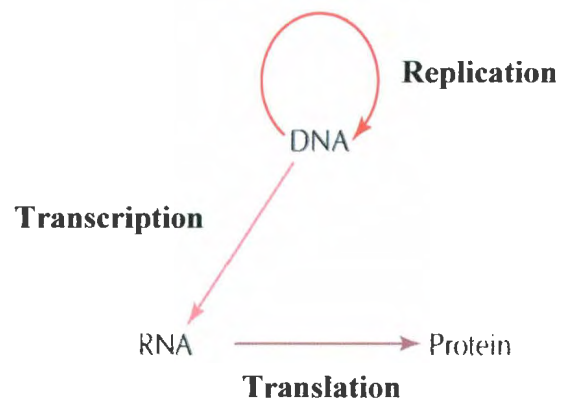
Common target genes especially implicated in this genetic dysregulation are oncogenes and tumour suppressor genes (Hanahan & Weinberg, 2000). Oncogenes, such as phosphatidylinositol-3-kinase (PI3K), Ras, and MYC, are genes that encode proteins that specifically control cell proliferation and differentiation (Croce, 2008). For example, the c-MYC oncoprotein, which is normally involved in stimulating cell growth, is overexpressed in many tumours and contributes to the tumour's growth potential (Croce, 2008). Conversely, tumour suppressor proteins, such as p53 and phosphatase and tensin homolog (PTEN), originate from genes that act to provide cellular restraints to prevent

abnormal growth and survival, as well as genomic instability. Loss or transcriptional repression of these genes results in the loss of growth control (Sherr, 2004).

As evident from information presented in this section, a principle outcome of genetic alteration is aberrant gene expression (Dupasquier & Quittau-Prévostel, 2009). Cells are a function of their gene expression, and consequently protein production, even the production of abnormally functioning proteins, is dictated by the regulation of this process. Therefore, understanding the molecular mechanisms at this level of the cellular regulatory network serves as a focal point for the characterization of the original factors initiating or inducing the development of diseases resulting from a lack of regulatory control, such as cancer.

## **1.2 Gene Expression**

Gene expression is the process by which the inherent genotype specified by the DNA sequence or genetic code of a cell is used as an instructional template for the production of proteins (reviewed in Orphanides & Reinberg, 2002). Explicitly described by the central dogma of molecular biology, the first step of gene expression is the production of pre-messenger ribonucleic acid (pre-mRNA) by a process called transcription (Crick, 1970). In eukaryotic cells, the resulting pre-mRNA molecule is then further processed into a mature messenger ribonucleic acid (mRNA) molecule through a series of splicing reactions and the addition of both a 5' cap and 3' polyadenylation tail. Mature mRNA molecules are then translated by the ribosomal complex into protein (reviewed in Sonenberg & Hinnebusch, 2009).



**Figure 1.3: The Central Dogma of molecular biology**

This figure, adapted from Crick, 1970, illustrates the central dogma of molecular biology which describes the residue by residue transfer of sequential sequence information whereby DNA can be copied to DNA through a process known as DNA replication or it can be copied into mRNA during transcription. Proteins can be then subsequently synthesized using the information in mRNA as a template during translation (Crick, 1970).



Organisms are essentially a function of their protein makeup, or phenotype. As gene expression is the manner by which proteins are generated, this phenomenon is of fundamental importance to the differentiation and development of all living things. The distinct and specific timing of gene expression initiation and the extent to which a particular gene is being expressed dictates how cells can differentially develop into separate cellular lineages, even though every cell in a living organism harbours the exact same genomic sequence (Emerson, 2002).

There are essentially three principal stages involved in the process of gene expression (Orphanides & Reinberg, 2002). These stages, some of which are briefly addressed in the following sections, are as follows:

1. Transcription
2. Post transcriptional processing
3. Translation

### **1.2.1 Transcription**

Transcription, or mRNA synthesis, is the generation of a complementary ribonucleic acid (RNA) molecule using the DNA sequence of the gene of interest as an instructional code or template (Deutschman, 2005). There are various steps in this particular process. The first is the transcription initiation stage which initially involves the binding of *trans*-acting factors, such as sequence specific and/ or basal transcription factors (TFs) to the promoter region (Boeger *et al.*, 2005; Deutschman, 2005). Factor binding then enables the recruitment and complex coordination of the necessary

transcription machinery molecules making up the pre-initiation complex (PIC) to the transcription start site (TSS) (Butler & Kadonaga, 2007; Davuluri *et al.*, 2008; Heintzman & Ren, 2007; Maston *et al.*, 2006; Morse, 2008). In most cases, it is this event that constitutes the rate-limiting step in transcription (Maston *et al.*, 2006). This is then followed by the elongation step by RNA polymerase II, which continues until a termination signal is encountered and both the RNA polymerase and the newly synthesized pre-mRNA are released (Deutschman, 2005; Morse, 2008).

#### **1.2.1.1 Promoter Regions**

Promoter regions are sequences of DNA that lie immediately upstream of the TSS, which is defined as the 5' end of the mRNA of a given gene and is the site within a gene where transcription begins (Landry *et al.*, 2003; Werner, 2003). These promoter regions are key elements involved in transcription as they are specifically recognized by the transcription initiation machinery. The binding of this machinery to the promoter recruits the RNA polymerase II complex to the site of transcription initiation, thereby facilitating elongation and mRNA synthesis (Carninci *et al.*, 2006; Heintzman & Ren, 2007). Typically, the term 'promoter' refers to the 'core promoter' and its adjacent sequences. The core promoter immediately surrounds the TSS and comprises 70–80 base pairs (bp) that contain sequence elements sufficient for recognition by the basal transcription machinery (Heintzman & Ren, 2007; Smale & Kadonaga, 2002). The 'proximal promoter' includes the region extending upstream of the core promoter (generally approximately 250 bp from the TSS) (Smale & Kadonaga, 2002). In mammals,



promoter regions are generally classified under one of two categories, either a conserved TATA box- enriched promoter, or a CpG rich promoter (Carninci *et al.*, 2006).

The TATA box is a regulatory *cis*-element with a consensus sequence of TATA\_A/T\_A\_A/T\_ and is typically located about 25–31 bp upstream of the TSS (Cao *et al.*, 2008). This regulatory *cis*-element tethers the PIC to DNA, thereby eliciting transcription initiation via its interaction with the transcription factor IID (TFIID) component, the TATA binding protein (TBP) (Hume, 2008). TATA box-enriched promoters, which can also be referred to as “sharp” promoters, initiate at a well-defined site and have very precise positional preference to the TSS. They are associated with tissue-specific gene expression and are highly conserved across species (Hume, 2008). These promoters, though, comprise only a small fraction (10–20%) of overall promoters. It is also important to note that a promoter may be classified as this type even if they do not contain a consensus TATA box (Hume, 2008).

Conversely, the other type of promoter, CpG rich promoters, constitutes the majority of eukaryotic promoters and is associated with ubiquitously expressed genes (Huang *et al.*, 2009). These promoters, which are also described as “broad” promoters, are more flexible with regards to initiation sites and do not harbour a single defined transcription start site (Carninci *et al.*, 2006; Levine & Tjian, 2003). Instead, the initiation of transcription at these promoters takes place throughout a cluster of nucleotide positions spread over an area of 50–100 bp (Carninci *et al.* 2005; Carninci *et al.*, 2006; Frith *et al.*, 2008). Frith *et al.*, 2008 attest that these specific promoters can therefore be accurately



described as a distribution of initiation site events on a stretch of nucleotides (Frith *et al.*, 2008).

### **1.2.2 RNA Splicing**

Various stages are involved in post-transcriptional processing; RNA splicing being one of the principle steps. RNA splicing involves the concomitant joining of exons (regions of mRNA transcript that encode the amino acid (aa) sequence of the protein) through the removal of non-coding sections, or introns, in the pre-mRNA (Matlin *et al.*, 2005). This process is controlled and orchestrated by the spliceosome, which is a macromolecular ribonucleoprotein complex that assembles on pre-mRNA (reviewed in Black, 2003). Consensus sequences at the end of introns, called splice sites, guide the assembly of the spliceosome and subsequent intron removal.

Ninety-eight percent of the human genome consists of intronic sequence and thus the removal of introns constitutes a central mode of genetic regulation. The timing and manner to which RNA splicing is employed affects gene expression, as alternate pre-mRNA splicing allows individual genes to express multiple mRNAs. This phenomenon serves as a key mechanism enabling differential protein production, as the generation of multiple protein isoforms with diverse functions can result from a single gene (Black, 2003; Matlin *et al.*, 2005; Venables, 2009). Production of variable mRNAs through alternate pre-mRNA splicing arises due to the incorporation of differing exons in mature mRNA transcripts and/ or the retention and/ or facultative usage of varying introns

(Venables, 2009). Consequently, such alternate inclusion of exons and introns can result in the incorporation of additional amino acids or can even shift the open reading frame at the level of translation thereby specifying diverse amino acid sequences in the resulting protein isoforms (reviewed in Black, 2003 and Venables, 2009; refer to section 1.3.1.2 for further discussion of alternate RNA splicing and promoter usage).

### **1.3 Regulation of Gene Expression**

The complexity of gene expression regulation is reflected by the multitude of intersecting elements, factors, and signal transduction pathways that have been demonstrated to influence an organism's phenotype (Emerson, 2002; Panning & Taatjes, 2008). For the purposes of this thesis and study, the main focus will be transcriptional regulation.

#### **1.3.1 Regulation of Transcription**

Cellular transcriptomes, the set of all RNA molecules produced in a particular cell type at a specific time, are essentially dictated by the interactions of *cis*-acting regulatory elements (sequences inherent in the genomic sequence, generally located within promoter regions) with *trans*-acting factors (Hu *et al.*, 2008). Of the most relevant *trans*-acting factors are transcription factors (TFs). TFs are generally categorized into two distinct groups: 1) the basal TFs, which are ubiquitous and recruit the RNA polymerase II multi-protein complex to the minimal or core promoter, and 2) gene-specific TFs that activate



or repress basal transcription by binding to regulatory DNA sequences (Morse, 2007; Villard, 2004). Each eukaryotic genome encodes between several hundred and several thousand TFs that work to either positively or negatively affect the rate of transcription (as activators and repressors, respectively) (O Barrera & Ren, 2006; Villard, 2004).

As the regulatory network expands beyond the transcription factor-DNA interactions to include molecules involved in the regulation of TFs specifically, the nexus of possible interconnections and interactions between many additional regulatory signalling molecules becomes even more extensive (Levine & Tjian, 2003). Consequently, transcription is also further controlled by: coregulators, cofactors, enhancers and silencers present in the promoter regions (Szutorisz *et al.*, 2005), chromatin condensation (Fry & Peterson, 2002), DNA methylation and acetylation via histone modifiers (Morse, 2007), RNA polymerase complexes, and even multiple promoter usage (Kleinjan & Lettice, 2008; O Barrera & Ren, 2006).

Coregulators, which pertain more specifically to aspects involved in this current study, are master regulator molecules that control transcription by tethering to or interacting with TFs, by modulating cytoplasmic signalling cascades, and by mediating posttranslational modifications of histones at promoter regions (Kininis & Kraus, 2008; Lonard & O'Malley, 2007; McKenna & O'Malley, 2002; O'Malley, 2008). Coactivators augment transcription by remodelling and altering chromatin conformation, nucleosome structure, and position, by modifying core histone proteins, as well as by bridging receptor complexes to the basal transcriptional machinery (Hall & McDonnell, 2004; Heintzman & Ren, 2007; Naar *et al.*, 2001; Smith & O'Malley, 2004). Conversely,



corepressors repress transcription through recruitment of histone deacetylases (HDAC), competition for coactivator binding, and recruitment of additional corepressors (Smith & O'Malley, 2004).

#### **1.3.1.1 Promoter Regions & the Regulation of Transcription**

Promoters intrinsically contain specific sequences of DNA that act as transcription factor-specific regulatory motifs, and are referred to as transcription factor binding sites (TFBS) (Cao *et al.*, 2008). These *cis*-elements facilitate the interaction and tethering of the promoter region DNA with TFs and other known enzymes involved in RNA polymerase recruitment. The pertinent interplay between these *cis*-regulatory elements and *trans*-acting TFs in specific cellular environments serves to coordinate transcription in time and space, as the promoter being recognized by the TFs and the time point of recognition are what essentially drive the initiation of transcription (Kadonaga, 2004; Landry *et al.*, 2003; Lemon & Tjian, 2000; Levine & Tjian, 2003; Maston, *et al.*, 2006; Mees *et al.*, 2009; Sakakibara *et al.*, 2007; Werner, 2003). Promoters furthermore incorporate environmental influences in the regulation of transcription as the condition of the cell will dictate the type of TFs that are available to recognize and bind to promoter region specific *cis*-elements.

#### **1.3.1.2 Alternate Promoter Usage**

Early theories regarding the regulation of transcription have been challenged and made more complex following the discovery that a single gene can be under the

regulation of multiple separate promoters. The alternate promoter regions of a gene have their own specific TSSs upstream of distinct alternatively encoded exons (Landry *et al.*, 2003). These additional promoters can contain unique *cis*-elements and therefore can each be specifically activated or repressed by distinct TFs, adding to the multifarious intricacies of transcript production, as well as to the spatial and temporal diversity in the expression patterns for a single gene (Davuluri *et al.*, 2006).

It was previously suggested that 18-20% of protein encoding genes use alternate promoters (Carninci *et al.*, 2006; Landry *et al.*, 2003). However, Carninci *et al.*, 2006, using a whole genome cap analysis of gene expression (CAGE) approach with sequence tags corresponding to several hundred thousand TSSs in the mouse and human genome, showed that differentially regulated alternative TSSs are a much more common feature than previously thought, with 58% of protein coding genes using two or more alternate promoters (Carninci *et al.*, 2006). Since this study, other whole genome studies have continued to illustrate the prevalence of gene expression regulation through alternate promoter usage, showing that more than 60% of protein coding genes are transcribed from multiple promoters and that these alternate promoters can span up to thousands of kilobase pairs (Baek *et al.*, 2007; Davuluri *et al.*, 2006; Kimura *et al.*, 2006).

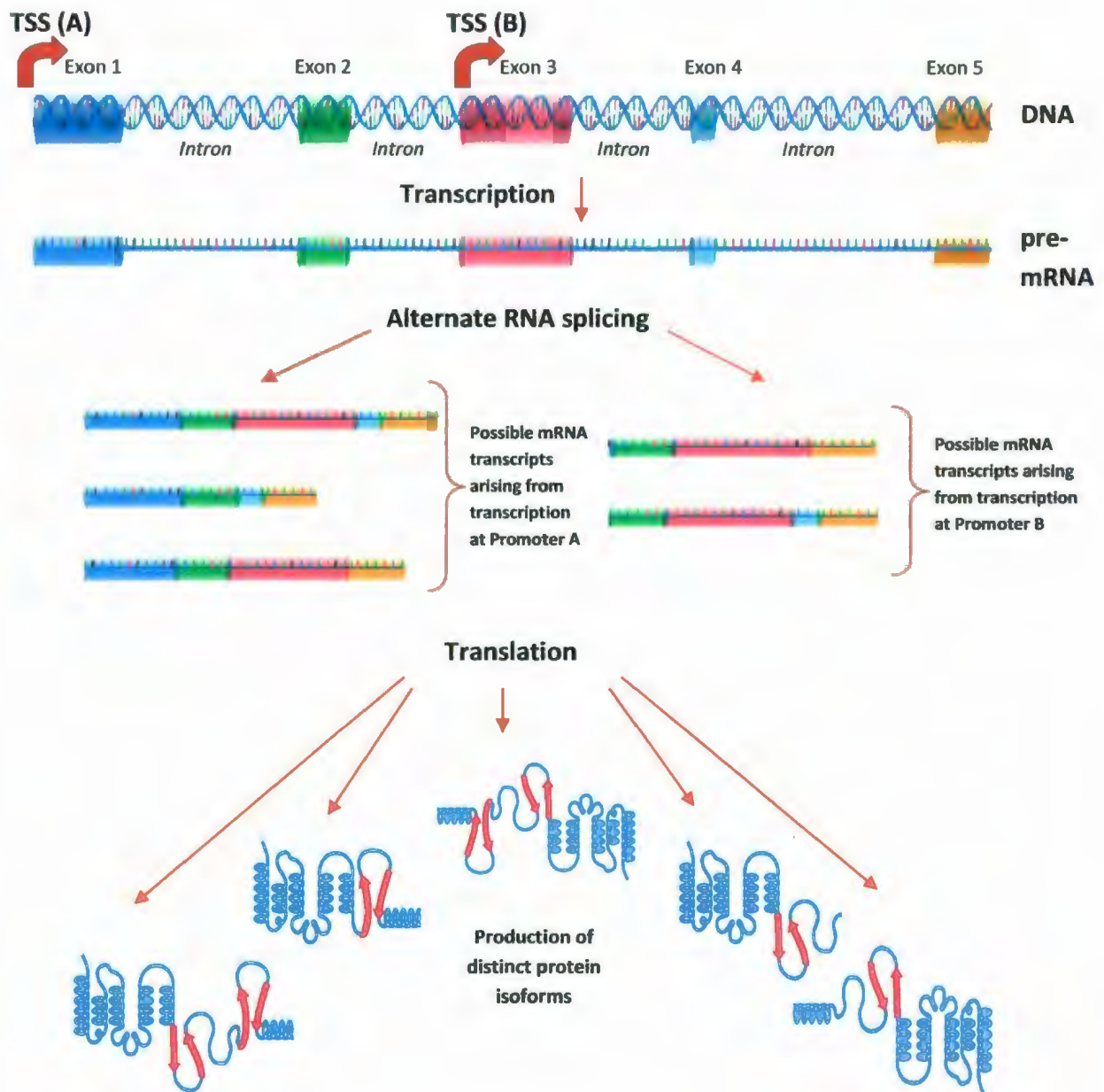
The nature of the promoter structure and the cellular environment dictate the usage of one promoter over the other (Zhang *et al.*, 2004). Cellular environmental influences, which vary from cell to cell and tissue to tissue, can include multiple TFs from distinct signalling pathways, regional epigenetic events (such as DNA methylation, histone modification, and chromatin remodelling), and even the participation of *cis*-



regulatory elements upstream of the promoter region (Davuluri *et al.*, 2006; Huang *et al.*, 2009). For example, the human reduced foliate carrier gene is expressed at different levels throughout the body due to differential regulation of its two promoters (for further examples of genes affected by alternate promoter usage refer to: Andres *et al.*, 2007; Dehma & Bonham, 2004; Hebden *et al.*, 2000; Koenigsberger *et al.*, 2000; Logette *et al.*, 2003; Manc *et al.*, 2005; Whetstine *et al.*, 2002).

There exist numerous consequences to alternate and/ or multiple promoter usage. One consequence, which has recently attracted considerable attention, is the role of alternate promoter usage in the regulation of developmental stage- specific and/ or tissue-specific expression. Many research groups have found that developmental stage-specific or tissue-specific expression of many genes, including receptor genes, are regulated by alternative promoters (Bharti *et al.*, 2008; Kakizawa *et al.*, 2007; Puomila *et al.*, 2007; Sehgal *et al.*, 2008; Turner *et al.*, 2006; Zhang *et al.*, 2004). Additionally, alternate promoter usage results in the generation of new exon combinations when transcribing from a single gene (Davuluri *et al.* 2006; Xin *et al.*, 2008). The differing mRNA transcripts arise from the incorporation of new exons associated with each respective promoter (refer to Figure 1.4). This incorporation leads to the production of variant 5' ends that then serve as templates for the generation of distinct N-terminal end domains in the subsequently translated proteins. The nature of the new alternatively incorporated exons to the 5' transcript ends will dictate whether the transcripts variants a) encode identical proteins with similar function or b) encode different, novel protein isoforms with diverse and even possibly antagonistic functional activities (Davuluri *et al.*, 2006).





**Figure 1.4: Alternate RNA splicing and promoter usage**

This figure illustrates the process of how one gene can encode varying protein isoforms through both alternate RNA splicing and alternate promoter usage. The hypothetical gene depicted here has two promoters (A & B) with distinct transcription start sites. The exons are marked and represented by altering colors. As demonstrated by the possible mature mRNA transcripts, both differential RNA splicing and transcriptional initiation at distinct promoters influences the diversity of transcript production, which can ultimately lead to the generation of distinct protein isoforms following translation.

The production of distinct protein isoforms that exhibit varying activity and/ or functions is mainly due to the fact that many alternate promoters are usually separated by one or more exons encoding important functional domains (Davuluri *et al.*, 2006). Protein interaction motifs and subcellular localization signals located at the N- terminus, such as secretory signal peptides and mitochondrial targeting signals, are just some examples of the possible functional domains that could potentially be encoded by variant transcripts. Furthermore, distinct protein isoforms can be generated following either a variation in splicing, or a switch in the open reading frame caused by the inclusion of alternate exons at the 5' transcript end (Landry *et al.*, 2003).

It is also worth noting that mammalian genes commonly employ multiple promoters to regulate translation as well (Landry *et al.*, 2003). One of the most influential processes controlling translation efficiency is the modulation of mRNA stability, a cellular process that is highly affected by the 5' untranslated regions (UTRs) of the mRNA (Smith, 2008). Consequently, as alternate promoter usage can dictate the composition of the 5' transcript ends, it highly affects the structure of 5'UTRs. It has been estimated that 10–18% of genes express alternative 5' UTRs by using multiple promoters (Trinklein *et al.*, 2003; Zhang *et al.*, 2004). Alternate promoter usage thus also produces variant UTRs along with variant transcripts. Moreover, the generation of variant 5' UTRs mediates tissue- and/ or cell-specific expression, as well as temporal-and/or developmental-specific expression, because different UTRs can contain distinct regulatory motifs that specifically control translation efficiency. Several past reports have

described genes that have alternate UTRs under a cell-type or tissue-specific manner (Hughes, 2006; Hughes & Brady, 2005; Zhang *et al.*, 2004).

## **1.4 Consequences of the Dysregulation of Gene Expression: A Focus on Carcinogenesis**

### **1.4.1 Abnormal Gene Expression & Cancer Development**

Dysregulation of transcription, and subsequent dysregulated gene expression, contributes to both cancer initiation and persistence (Mees *et al.*, 2009; Villard, 2004). Furthermore, signals involved in promoting cancer are principally controlled by the amplified expression of oncogenes and/ or the loss of tumour suppressor expression (Croce, 2008; Mees *et al.* 2009). All principle stages of gene expression have been implicated in carcinogenesis when improperly activated (Smith, 2008; Stoneley, 2003; Venables, 2009). For the purposes of this current thesis and study though, the following sections will focus on how alternate promoter usage and both aberrant TF and coregulator signalling contribute to the oncogenic state.

### **1.4.2 Abnormal Promoter Usage & Cancer Development**

Comparable to most key regulatory processes in the cell, the dysregulation of alternate promoter usage has been implicated in the development of various diseases (Davuluri *et al.*, 2006; Hughes, 2006; Nakanishi *et al.*, 2006; Pedersen *et al.*, 2006). Recently, more and more evidence is surfacing that supports a pivotal link between aberrant alternate promoter usage and cancer, as several oncogenes and tumour



suppressor genes contain multiple promoters that become dysregulated throughout tumourigenesis. In some of these cases, an abnormal preference in promoter usage has even triggered the initial formation of neoplasia (Davuluri *et al.*, 2006).

One specific example that demonstrates how aberrant alternate promoter usage can result in cancer development is that of the CYP19 gene, which itself is under the regulation of eight separate promoters (Agarwal *et al.*, 1996; Chen *et al.*, 2009). CYP19 encodes the cytochrome P450 enzyme aromatase, which is required for the biosynthesis of estrogen (reviewed in Chen *et al.*, 2009). Disproportionately high expression of this gene has been documented to play an important role in the development and progression of breast cancer (Agarwal *et al.*, 1996; Bulun *et al.*, 2009; Chen *et al.*, 2009). Studies performed by Agarwal *et al.*, 1996 revealed that alternative promoter usage is a major mechanism that mediates abnormal increased aromatase expression in breast cancer, as breast cancer patients more abundantly expressed transcripts deriving from a promoter that, under normal conditions, is quiescent (Agarwal *et al.*, 1996; Bulun *et al.*, 2009). Furthermore, 20% of known aromatase expression in breast cancer has been found to originate from the aberrant activation of an additional distally located promoter of CYP19, the I.7 promoter (Bulun *et al.*, 2009; Chen *et al.*, 2009).

#### **1.4.3 Aberrant Transcription Factor & Coregulator Signalling in Cancer Development**

As TFs essentially dictate when a particular gene will be expressed, it is no surprise that TF dysregulation, either the over-activation or repression of their signalling, has been implicated in carcinogenesis (Darnell, 2002; Mees *et al.*, 2009).

As illustrated in Table 1.1, Mees *et al.*, 2009 outlined various transcription factors in accordance to their influence on the progression of the original six hallmarks of cancer. Moreover, products of oncogenes and tumour suppressor genes are often potent transcription factors (Mees *et al.*, 2009). A number of developmental stages, as well as cell type differentiation, are under the control of oncogenic transcription factors. Moreover, the loss of anti-oncogene TFs encoded by tumour suppressor genes has been known to result in carcinogenesis (Villard, 2004; Mees *et al.*, 2009). For example, germ-line mutations in TFs with tumour suppressor function have been shown to be responsible for various hereditary cancers, such as Li-Fraumeni syndrome, Wilms' tumour, and retinoblastoma (Villard, 2004). (Refer to Darnell, 2002; Mees *et al.*, 2009 for more an in-depth review on transcription factors and cancer).

**Table 1.1 Hallmark ascriptions of oncogenic transcription factors**

**Table 1** Hallmark ascription<sup>a</sup> of oncogenic transcription factors

Transcription factor	Hallmark traits					
	Self-sufficiency in growth signals	Insensitivity to growth-inhibitory signal	Evasion of programmed cell death	Limitless replicative potential	Sustained angiogenesis	Tissue invasion and metastasis
AP-1	X		X		X	X
AR	X					
ATF-1						X
BRN-3b		X				
C/EBP $\alpha$		X				
CREB			X		X	
E2F-1				X		
ETS-1					X	X
EWS/ETS				X		
FOXO3a			X			
HIF-1 $\alpha$ /HIF-1 $\beta$ (ARNT)					X	X
Myc	X				X	X
NF- $\kappa$ B	X		X		X	X
PEA3						X
RAR $\alpha$	X					
RB1				X		
SP-1					X	
STAT3		X	X		X	
STAT5		X	X			
TP53			X			

➤ From Mees *et al.*, 2009<sup>3</sup>.

<sup>3</sup> License agreement between JC and Elsevier provided by Copyright Clearance Center. License Number: 2391090899689; License date: March 16, 2010



Coregulators, which interact with and regulate TFs, have also been implicated in cancer. Numerous studies have revealed that the expression of selected coactivators is associated with tumourigenesis and/ or progression in both breast and prostate cancer, and further sensitize the tumours to estrogens and growth factors (Smith & O'Malley, 2004). Moreover, high levels of corepressor nuclear receptor corepressor 2 (NCOR2) and Silencing Mediator for Retinoid and Thyroid Hormone Receptors (SMRT) expression has been shown to be associated with poor patient outcome in breast cancer, independent of other prognostic factors. Conversely, other steroid receptor coactivator expression is generally associated with good prognosis (Green *et al.*, 2008). Such evidence has furthermore demonstrated that an imbalance of coactivators and corepressors may contribute to the acquisition of resistance to endocrine therapy in these aforementioned cancers (Schiff *et al.*, 2003; Gururaj *et al.* 2006).

### 1.5 Mesoderm Induction Early Response 1 (*mier1*)

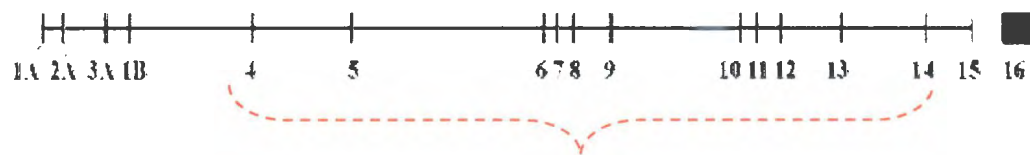
Mesoderm induction early response 1 (*mier1*) is a gene encoding a potent transcriptional regulator protein. This gene was first discovered in *Xenopus laevis* in response to fibroblast growth factor signalling (Paterno *et al.*, 1997). Fibroblast growth factor is highly implicated in cell differentiation, mitogenesis, motility and angiogenesis (Grose & Dickson, 2005). Capable of inducing embryonic cells to differentiate into mesodermal tissues, this particular growth factor has also been observed to cause phenotypic transformation in various cell lines when over-expressed (reviewed in Grose & Dickson, 2005; Sasada *et al.*, 1988). This succinctly demonstrates the potent regulatory function of this growth factor and consequently the regulatory potential of responding genes, such as *mier1*, which serve to further propagate this growth factor's signal throughout the cell at differing time points of development and during varying metabolic processes.

Concordantly, after being characterized as an immediate early gene and target of the fibroblast growth factor signal transduction pathway (Paterno *et al.*, 1997) in *Xenopus laevis* embryo explants, and later being cloned and characterized in humans using a human testes complementary DNA (cDNA) library (Paterno *et al.*, 1998), *mier1* has ably demonstrated its puissant regulatory potential. Depending on the cellular context, the *mier1* protein product, MIER1, can act as either a repressor or activator of transcription, functions of which are dictated by the underlying gene and protein structure.

### 1.5.1 *mier1* Gene, Transcripts, & Protein Isoform Structure

*mier1* is a complexly organized 63 kilobase pair (kb) gene that is located on chromosome 1 at position p31.2 (Paterno *et al.*, 2002). The intricacy of its structural organization stems from the fact that it contains 17 different exons which, following exon skipping, differential polyadenylation, and both facultative intron and alternate promoter usage, give rise to 12 different mRNA transcripts that can ultimately be translated into 6 distinct protein isoforms (refer to Figure 1.5 for a gene structure schematic) (Paterno *et al.*, 2002). This vast assortment of protein isoform production derives from the aforementioned transcriptional and post-transcriptional processes at either the 5' or the 3' end of *mier1* and its transcripts.





**Figure 1.5: *mier1* Gene Structure**

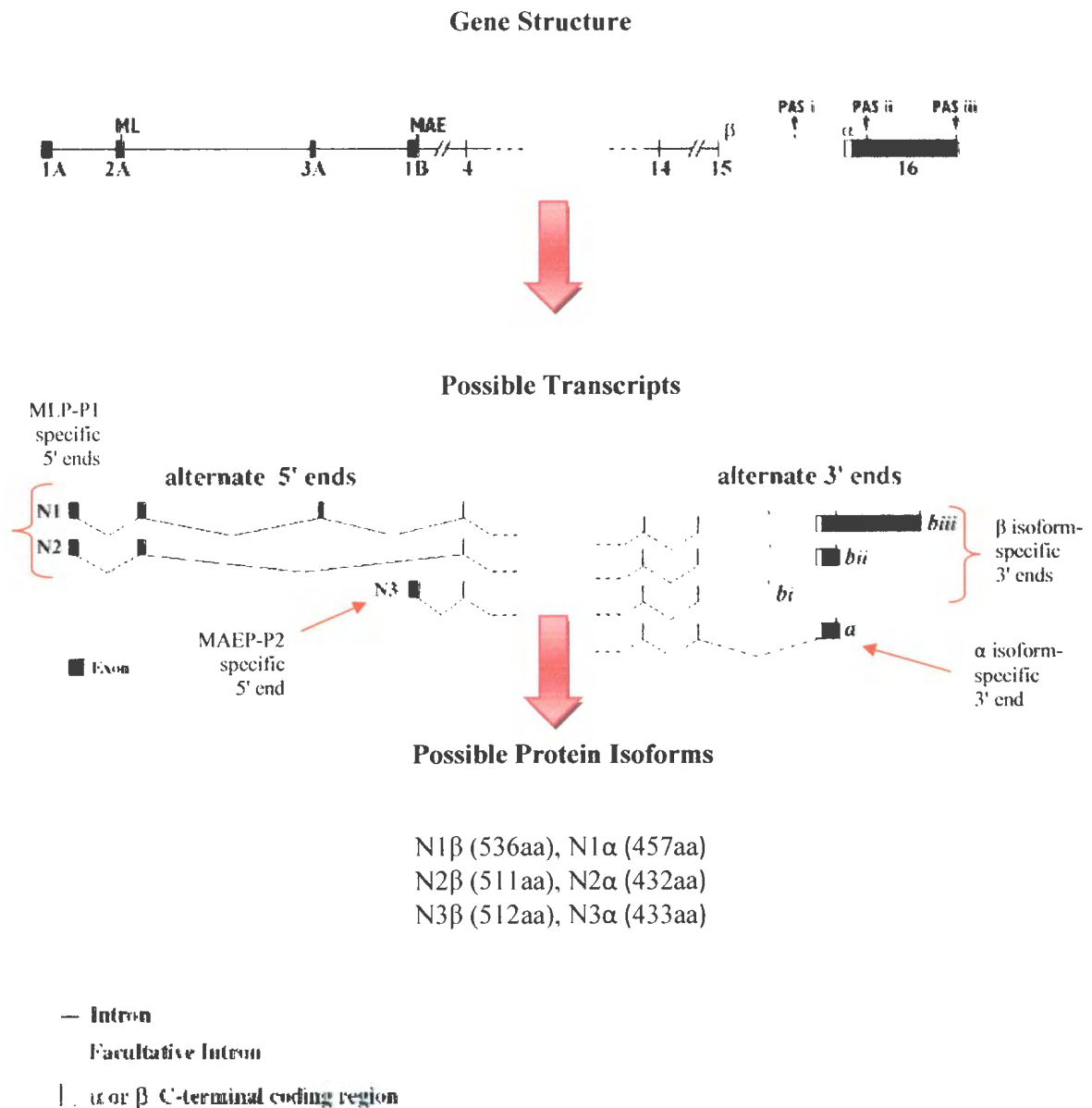
This figure is adapted from Paterno *et al.*, 2002. Exons are numbered and the common region between all *mier1* protein isoforms is indicated by the red dotted bracket.

With respect to the 3' end of *mier1*, there are three possible polyadenylation sites which, depending on the specific site targeted, dictate the structure of the final transcript (Paterno *et al.*, 2002). Figure 1.6 depicts the location and outcome of the polyadenylation of each site, which are marked polyadenylation site (PAS)i, PASii, and PASiii respectively. Differential polyadenylation does not, however, have an effect on the overall nature of the C-terminal protein domain, as the C-terminal coding region is not altered. This is unlike the most notable variation in the 3' transcript ends, which is facultative intron usage. The incorporation of the final intron, intron 15, as a facultative intron provides the final protein isoform with a distinct coding region that results in a 102 amino acid C-terminal domain. This facultative intron incorporation is the determining factor that demarks the protein as the beta ( $\beta$ ) isoform (Paterno *et al.*, 2002). A protein harbouring this facultative intron domain is referred to as MIER1 $\beta$ , regardless of the location of polyadenylation and the N-terminal end composition. Proteins comprising the second C-terminal domain, termed alpha ( $\alpha$ ), have been translated from transcripts that do not use intron 15 as a facultative intron and therefore contain only 23 distinct amino acids in the C-terminal domain. All proteins translated from transcripts lacking the facultative intron are termed MIER1 $\alpha$ , regardless of N-terminal composition and variable polyadenylation (Paterno *et al.*, 2002). Exons 4-15 are constitutive exons that encode 410 residues, which make up the common internal or core sequence of the *mier1*-derived protein isoforms (Paterno *et al.*, 2002). These exons do not contribute to transcript diversity; they do, however, contain important regulatory motifs implicated in MIER1's cellular function (discussed further in section 1.5.3.).

Another element affecting diverse transcript production is alternate promoter usage. Depending on which of the *mier1* promoters becomes activated, either MLP-P1 or MAEP-P2, will dictate the organization of the 5' ends of the transcripts, and ultimately the N-terminal end of the respective protein isoform. Figure 1.6 delineates the possible 5' transcript ends produced by each separate promoter. Transcriptional activation at the MLP-P1 promoter will produce N-terminal ends "N1" and "N2", while activation at the MAEP-P2 promoter will produce N-terminal end "N3". The difference in the two N-terminal ends generated by MLP-P1 are due to skipped exon usage, in which "N1" includes the skipped exon 3A (74bp in length), which is inserted after the first amino acid (M) encoded by exon 2A (Paterno *et al.*, 2002).

As further illustrated by Figure 1.6, the twelve mRNA transcripts arise through the mixing and matching of the four possible 3' ends with the three possible 5' ends. Subsequently, the six possible protein isoforms are generated by individually combining each of the three possible N-terminal ends to either the  $\alpha$  or the  $\beta$  isoform specific C-terminal end (Fig 1.6).





**Figure 1.6 : *mier1* Transcript & Protein Isoform Combinations**

This figure is adapted from Paterno *et al.*, 2002. The 'gene structure' section can be used as a reference to the exon numbers. It also illustrates the alternate AUG start sites (ML, MAE) for both of *mier1*'s promoters, whereby transcription initiation at the first promoter, MLP-P1, results in N2 and the amino acids ML. Although not depicted in the above figure, N, which includes the alternate exon 3A results in the amino acid sequence MFMFNWFTDCLWTLFLSNYQ. Transcription initiation at the second promoter, MAEP-P2, results in N3 and the amino acids MAE. Furthermore, the alternate poly A sites are marked by PASi, PASii, and PASiii. Below the "gene structure" diagram the alternate transcripts and their possible respective 5' and 3' ends are depicted. N1, N2, N3 represent the alternate N-termini and biii, bii, bi represent the structural outcome of the alternate poly A sites respectively. The  $\alpha$  and  $\beta$  specific 3' ends are also indicated.

### 1.5.2 *mier1* Promoters

Recently, the role of alternate promoter usage in gene regulation has become more apparent, as this regulatory phenomenon not only dictates tissue-specific expression, but has also been observed to be abnormally employed in various cancers (Agarwal *et al.*, 1996; Bulun *et al.*, 2009; Dehma & Bonham, 2004; Duan *et al.*, 2008; Li *et al.*, 2009; Miyazaki *et al.*, 2009; Renaud *et al.*, 2007; Benz *et al.*, 2006). As described in section 1.3.1.2, an important aspect contributing to the diversity of the possible *mier1* mRNA transcripts is the alternate usage of either of the gene's two promoters.

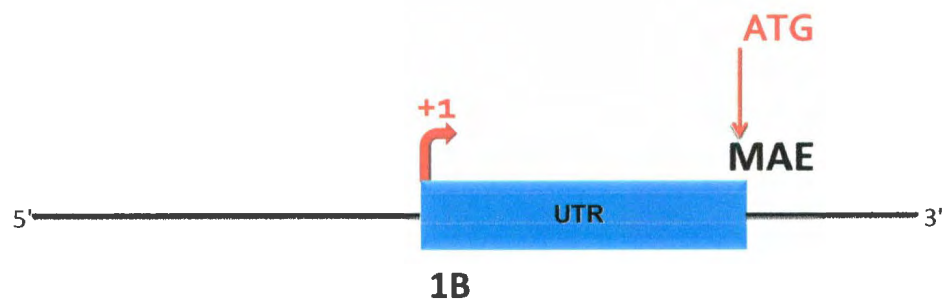
#### 1.5.2.1 MAEP-P2

*mier1*'s second promoter, MAEP-P2, which is named based on the three distinct N-terminal amino acids it encodes, was the first *mier1* promoter to be cloned and characterized. It is situated approximately 5.3kb downstream of the first promoter (Ding *et al.*, 2004) (Refer to *National Center for Biotechnology Information* (NCBI) Reference Sequence: NT\_032977.8). Transcriptional initiation at this promoter results in the inclusion of exon 1B and the generation of the 'N3' N-terminal domain thereby resulting in the translation of the three amino acids MAE (methionine, alanine, and glutamic acid), which are incorporated at the N-terminal end of the protein (refer to Figure 1.6 for the structure of the transcripts and possible N-terminal ends). It is also important to note that MAEP-P2 specific transcripts contain completely distinct 5' UTRs compared to transcripts produced from *mier1*'s first promoter MLP-P1, due to the incorporation of exon 1B. As briefly alluded to in section 1.3.1.2, the presence of

altering UTRs in promoter specific transcripts can potentially function to regulate gene expression.

Full-length 5' ends of *mier1* were furthermore amplified using the 5' Rapid Amplification of cDNA Ends (RACE) PCR method, and following sequence analysis and *Basic Local Alignment Search Tool* (BLAST) comparison, the TSS of this promoter was found to reside 149bp upstream of the MAEP-P2 ATG translation start site (Ding, 2004; Ding *et al.*, 2004). The following figure illustrates the structure of this promoter.

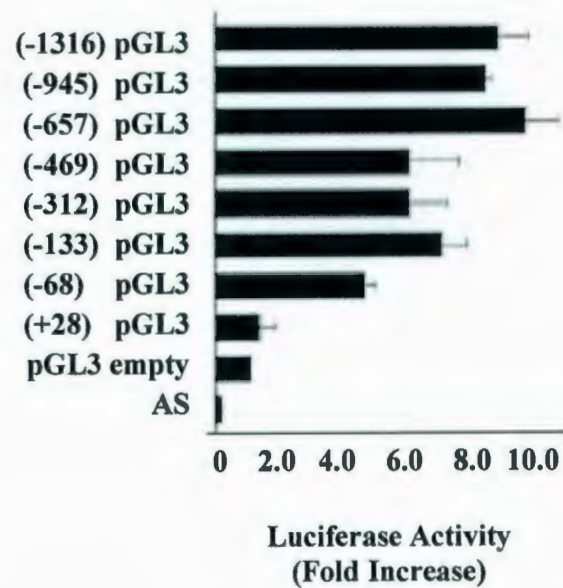




**Figure 1.7: MAEP-P2 promoter structure**

This diagram depicts the MAEP-P2 promoter region and the exon of *mier1* (exon 1B) that results following transcription at this specific promoter. Exon 1B, represented by the blue box, is 121 bp long and encodes the UTR as well as the first three amino acids of the protein (MAE). The transcription start site, represented by +1, begins 149 bp upstream of the ATG translation start site.

Further examination and computer-assisted sequence analysis of this promoter region revealed that MAEP-P2 lacks any identifiable TATA element, initiator sequence, and downstream promoter elements; the two latter of which are often found in TATA-less promoters (Ding, 2004; Ding *et al.*, 2004). It does, however, contain a CpG island located 389 nucleotides upstream of the TSS and extending until the start of translation (Ding, 2004) (Fig 1.9). Previous studies have also mapped out the location of the minimal promoter (the minimal sequence that is necessary for promoter activity). Following a deletion series analysis using a luciferase reporter, the minimal MAEP-P2 promoter was observed to reside 68bp upstream of the TSS. Figure 1.8 shows the different luciferase promoter constructs used, and their relative promoter activity following transient transfection into HeLa cells.



**Figure 1.8: Characterization of MAEP-P2 promoter activity**

This diagram, adapted from Ding *et al.*, 2004, depicts the relative luciferase activity following transfection of HeLa cells with various MAEP-P2 luciferase reporter gene deletion constructs (Ding, 2004; Ding *et al.*, 2004). The counts in the diagram represent the sequences cloned into the reporter gene construct and are with reference to the TSS. For the purposes of the current study, four of these particular constructs were employed: +28, -68, -312, and -1316 were used in this study. Refer to section 2.2.3, 3.4.1, and 4.6 for further details.



Furthermore, *in silico* analysis of this promoter region revealed several potential TFBSs, including multiple binding sites for the Specificity Protein 1 (Sp1) family (Ding *et al.*, 2004). Figure 1.9 illustrates the location of these putative TFBSs. The identified Sp1 binding sites were further investigated with regards to their role in regulating *mier1* transcription from MAEP-P2 by electrophoretic mobility shift assays (EMSAs) using a probe consisting of the MAEP-P2 minimal promoter sequence and a Sp1-specific antibody against HeLa cell nuclear extracts. Results demonstrated that Sp1 binds to the minimal promoter (Ding *et al.*, 2004). Furthermore, this interaction induced transcription at the MAEP-P2 promoter, as revealed from results following co-transfection of MAEP-P1 luciferase reporter gene deletion constructs and with Sp1 expression vectors (Ding *et al.*, 2004).

↓ CpG Island Begins

```

(-416/-565) caaggcataa ccgctttcaa agcacagttc ttgccgccgt tgcacacag caacatccgt
(-356/-505) ttttcagcaa atgcatttca aaaacgacct acatgtaaaa tattacgta Elf-1
(-296/-445) ctgtgtcaat gctggatgtg acleclttgg cactgttcc ERE tgacctctct gatccgtaaa
(-236/-385) caaccccctc cgccgggagt ggcggatcag ctggagccag cgaagcccc H4TF-2 cgcgcggttg
(-176/-325) cccatctect TEF-2 cccapaccca ccttgactcc TEF-2 gccccctccg ctcttccgg ggagggc gg
(-116/-265) ccgcggggcc TEF-2 gcgcgcgcgc cctgtctccg gcgcgtgctc gctgggtcttt tccctccagt
(-56/-205) ccagcccagc cggggcgccg cqagggggcg TEF-2 gagtcgggtg tggtgggcgc gctggcCGG +1
(+5/-145) CTCTGCGCG TTCCGCGCA GGCAGTGGCG GCGGGAGCGG CAGAGACGGC AGCGGCCGGA
(+65/-85) GTCCCGTTGC TGAGTCTCAC ATCCGGGTTC TGGCCGTGAC CCAGCTGCGG CCGCCGCGGA
(+125/-25) GATGTGACCC GGCAGTACGG CAAATATGC G

```

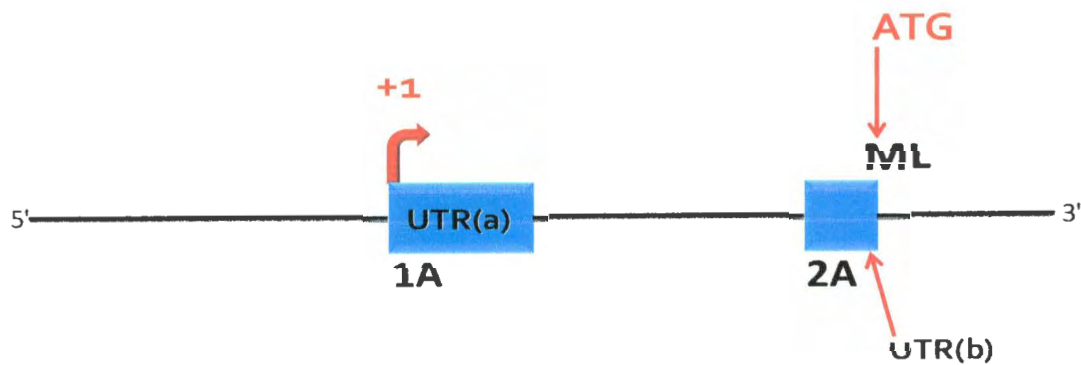
**Figure 1.9: MAEP-P2 promoter CpG Island sequence & putative transcription factor binding sites**

This schematic illustrates the nucleotide sequence of the MAEP-P2 promoter region. Sp1 sites are denoted by the navy blue boxes. Other transcription factor binding sites were also found following *in silico* analysis and are labelled, and the nucleotides constituting the binding sites are italicized with distinct colors. The TSS is represented by +1 and the beginning of the CpG island is specified by the red up-pointing arrow. Exon 1B encoded following activation at the MAEP-P2 promoter is denoted by the capitalized nucleotides.

### 1.5.2.2 MLP-P1

*mier1*'s most 5' promoter, MLP-P1, named based on the two distinct amino acids it encodes, had yet to be fully characterized prior to this study. Transcriptional initiation at this promoter results in the generation of two distinct N-terminal ends: N2, resulting from exon 1A and 2A, and N1 resulting from exon 1A, 2A, and a third normally skipped exon denoted exon 3A (refer to Figure 1.6 ) that lies 2.58 kb downstream of the ATG translation start (Paterno *et al.*, 2002). Exon 3A is 74bp in length and is inserted after the first amino acid (methionine) encoded by exon 2A. Inclusion of exon 3A results in the generation of the following distinct N-terminal end: MFMFNWFTDCLWTLFLSNYQ. Further characterization with regards to the minimal promoter region, putative TFBSs, and the location of the actual TSS has yet to be fully investigated. Figure 1.10 depicts the MLP-P1 promoter region structure, while Figure 1.11 illustrates the nucleotide sequence of the proximal promoter region.





**Figure 1.10: MLP-P1 promoter structure**

This diagram depicts the MLP-P1 promoter region and the two exons (represented by the blue boxes) of *mier1* that are incorporated into *mier1* transcripts following transcriptional activation at the MLP-P1 promoter. Exon 1A or UTR(a) is 137bp long, while exon 2A, which encodes for UTR(b) of MLP-P1 specific mRNA transcripts, also encodes for the first two amino acids (methionine and leucine) of the protein. Exon 2A is 101bp long. The putative transcription start site, or +1, is thought to be at the commencement of exon 1A and is 1279 bp upstream of the ATG translation start site, if the size of the intron is included. The intron, which is 1046bp long, resides between exon 1A and 2A. Without the intron included in the count, the TSS is 232bp upstream of the ATG translation start site. Although not depicted in this schematic, exon 3A lies 2.58kb downstream of the MLP-P1 translation start site.

```

(-149/-1528) gtggcgacca gctggggagt ggtgcaaccac cccctttttt agccgcctct gaagtcctctg
(-189/-1468) tacccccaag ctccctccgtt agcggtctcg gccgaggtct cgaatgttt gccggggtct
(-129/-1408) atggcgacgg tggagccctg gctcaacaag cggccgcgcg gtgggtctgg ggcacgaggt
(-69/-1348) cgaggaggag ggcggagggc gaggggaggg cagagggttg gtggagctgg aggaagctcc
Putative TSS ( + 1 ) / Exon 1A
(-9/-1288) qqacgacgaC TGAAGAAGG AGGCGGGCGG CCCGGGCTC AGGCGGCTCC CAGGCTCTGA
(+52/-1228) GTCTCCCGGC TGCAGGCGGA TGGATGGGGC TTCTTCAGGC GGTGGCGGCA GCAGCGAAGG
(+112/-1168) TGGCGGCGGC AGCAGCGGCA GCGGCT.....1046bp Intron.....
Exon 2A
(+1181/-95) ...ATGGTGT GGTGCTCGA TTCTCCAGT GCCTGGCTGA GTTTCGACG TGGTTAAGAA
(+1241/-38) CCAACTGGTT GAGGTTCAT GCAGACAAGA CGGATGTGAT GCTG
Translation Start Site

```

**Figure 1.11: MLP-P1 proximal promoter sequence**

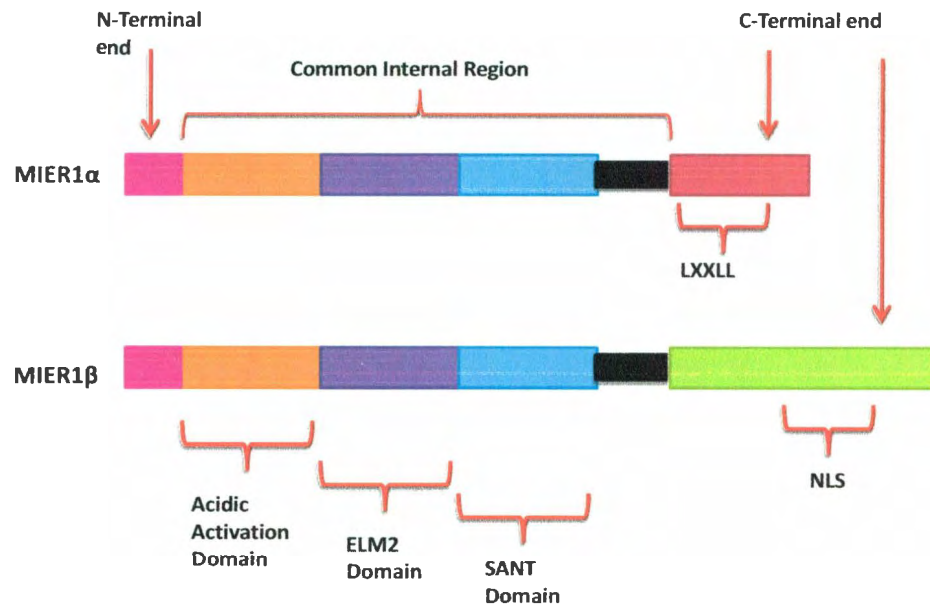
This figure illustrates the MLP-P1 promoter proximal region sequences. Uppercase letters represent the specific exons that are denoted by the boxes. The ATG translational start site, the putative TSS, and the 1046bp intron are also demarcated.

### 1.5.3 MIER1 Function

#### 1.5.3.1 MIER1 as a Transcriptional Activator

Primary elucidation of MIER1 function demonstrated its potential role as an activator of transcription through a series of transactivation assays using a CAT reporter gene vector (Paterno *et al.*, 1997). Various regions of the *Xenopus* ortholog of *mier1*, *xmier1*, which displays 91% overall similarity to the human ortholog at the amino acid level, were fused to the DNA binding domain of GAL4 and inserted into the chloramphenicol acetyltransferase (CAT) reporter gene vector (Paterno *et al.*, 1997; Paterno *et al.*, 1998). Results following transient transfection of these reporter gene vectors into NIH 3T3 cells illustrated XMIER1's ability to transactivate gene transcription as an increase in the level of transcription was observed following vector transfection (Paterno *et al.*, 1997). This activation arised from the presence of highly acidic regions or stretches that are clustered in the common region of the protein towards the N terminus (Fig 1.12) as the particular CAT reporter vectors harbouring this N-terminal region induced transcription 80-fold in this study (Paterno *et al.*, 1997). These specific types of acidic activation domain regions, of which there are three in human MIER1, are highly conserved between both the *Xenopus* and human *mier1* orthologs. Moreover, such acidic activation domains have previously been implicated as putative *trans*-activating domains involved in transcriptional activation, and have been found in various TFs (Blair *et al.*, 1994; Ko *et al.*, 2008; Melcher, 2000).





**Figure 1.12: Structure of *mier1* protein isoforms including protein functional domains**

This schematic represents the structure of the *mier1* isoforms with emphasis on the varying structural and functional domains. The common region shared by both isoforms is represented by the orange, purple, blue, and black boxes and consists of 410 amino acids. The acidic activation domains, of which there are three in humans, are represented by the orange box and consist of amino acids 29-37, 72-94, and 150-160 of the 410 amino acid common internal region. The ELM2 domain, depicted by the purple box, consists of amino acids 169-272 of the common region, while the SW13, ADA2, N-CoR, and TFIIB (SANT) domain, denoted by the blue box consists of amino acids 277-321. The  $\alpha$ -specific C-terminal end is made up of 23 amino acids and the LXXLL domain is situated between the 3<sup>rd</sup> and 7<sup>th</sup> amino acids of the C-terminal end. The  $\beta$ -specific C-terminal end consists of 102 amino acids whereby the nuclear localization signal (NLS) is located between the 64<sup>th</sup> and 70<sup>th</sup> amino acids of this C-terminal end.

### 1.5.3.2 MIER1 as a Transcriptional Repressor

Further functional studies of MIER1 have characterized this protein as not only having a role in gene transcription activation, but more frequently functioning as a repressor of gene expression (Ding *et al.*, 2003; Ding *et al.*, 2004; McCarthy *et al.*, 2008). Moreover, this repressive role is exerted through various mechanisms, each of which employ distinct protein functional domains that are inherent to either the central core/ common coding region of all *mier1* protein isoforms, or to the C-terminal domain of MIER1 $\alpha$  isoforms.

#### 1.5.3.2.1 Transcription Factor Tethering via the SANT domain

The signature SANT domain of *mier1* is situated approximately 277 amino acids downstream of the start of the common internal region contained in all protein isoforms. This particular domain exhibits 100% identity between human and *Xenopus laevis*, and following multiple genome base sequence analyses, also portrays 88% identity across nine different species, thereby highlighting its importance to MIER1 function (Thorne *et al.*, 2005). First identified in transcription factors SW13, ADA2, N-CoR, and TFIIB, the SANT domain is a protein motif that is implicated in DNA binding and protein-protein interactions, and is present exclusively in nuclear proteins that have an important role in the regulation of transcription (Aasland *et al.*, 1996; Ding *et al.*, 2004). Moreover this domain or motif is also known for its repressive regulatory function, as it was found to be present and implicated in the functional activity of the corepressors and subunits of histone deacetylases (HDACs), nuclear receptor corepressor (NCoR), and SMRT ( Boyer *et al.*, 2004).

Ding *et al.*, 2004 further supported the SANT domain's repressive regulatory potential in the context of MIER1 function (Ding *et al.*, 2004). An intact SANT domain was required for the physical interaction between the Sp1 transcription factor and MIER1, as demonstrated by glutathione s-transferase (GST)-pull down assays with various MIER1 deletion constructs (Ding *et al.*, 2004). Moreover, this particular study demonstrated that both isoforms of MIER1( $\alpha$  &  $\beta$ ), through a SANT domain mediated interaction with Sp1, displace Sp1 from its cognate binding sites and represses Sp1- activated transcription from its own minimal promoter (the MAEP-P2 minimal promoter) (Ding *et al.*, 2004). Furthermore, overexpression of MIER1 resulted in the loss of detectable Sp1 from this promoter (Ding *et al.*, 2004). Co-immunoprecipitation (co-IP) and chromatin immunoprecipitation (ChIP) analysis additionally demonstrated that not only do Sp1 and MIER1 physically associated with one another, but that they also associated with the chromatin at the *mier1* MAEP-P2 promoter endogenously (Ding *et al.*, 2004).

The ability of MIER1 to repress Sp1-mediated transcription clearly identifies a negative feedback loop regulating *mier1* isoforms and represents a novel mechanism for the negative regulation of Sp1 target promoters (Ding *et al.*, 2004). Sp1 itself is a TF from the family of Sp/Krüppel-like factors (KLF) family of TFs that bind to GC rich sequences, such as: GC boxes, CACCC boxes (also known as GT boxes), and basic transcription elements (collectively referred to as 'Sp1 sites') in gene promoter regions (Black *et al.*, 2001). Sp and KLF proteins cooperatively interact with one another and other transcription factors on GC-rich promoters to either activate or inhibit multiple and diverse classes of mammalian genes that play a critical role in regulating cellular homeostasis (Black *et al.*, 2001; Safe & Kim, 2004). Furthermore



Sp1-specific signalling has been shown to be induced following a number of varying cellular processes and stimuli such as growth stimulation, oncogene overexpression, and anti-metabolites (Black *et al.*, 2001). Thus, the nature of Sp1's role in the cell only further emphasizes the importance of the repressive regulatory potential exerted by MIER1.

#### **1.5.3.2.2 Recruitment of HDAC1 via the ELM2 Domain**

Another repressive regulatory function of MIER1 was attributed to its EGL-27 and Metastasis Associated Antigen 1 (MTA1) homology 2 (ELM2) domain, which is situated approximately 169 amino acids downstream of the start of the common internal region of *mier1* (Ding *et al.*, 2003- refer to Figure 1.12). ELM2 was first described in Egl-27, which is a *Caenorhabditis elegans* protein that plays a fundamental role in patterning and embryonic development (Ding *et al.*, 2003; Solari *et al.* 1999). This domain is often found N-terminally to a SANT- and a GATA-binding domain and thought to be involved in DNA-binding and/ or protein-protein interactions (Lakowski *et al.*, 2006). Moreover, the ELM2 domain and the SANT domain have been found to be present in several other important transcriptional repressors such as metastasis tumour antigen (MTA) (metastasis-associated family proteins), and CoREST (RE1 silencing transcription factor (REST) co-repressor) proteins (Wang *et al.*, 2008; Lakowski *et al.*, 2006).

Ding *et al.*, 2003, showed that MIER1 can function to repress transcription by recruiting a histone deacetylase 1 (HDAC1). HDACs, such as HDAC1, are chromatin remodelling enzymes that inherently function to inhibit transcription by decreasing the accessibility of essential transcription machinery to genomic DNA by increasing histone-DNA interaction the removal of acetyl groups. This consequently increases the histone's affinity for the DNA backbone, thereby making it less accessible.

Results obtained from transient transfection experiments showed that both isoforms significantly repressed expression of the G5tkCAT reporter plasmid in a dose-dependent manner (Ding *et al.*, 2003). Immunoprecipitation (IP) assays revealed that both MIER1 $\alpha$  and MIER1 $\beta$  induced transcriptional repression by recruiting HDAC1 *in vivo*. Moreover, addition of an inhibitor of HDAC1 activity partially relieved the MIER1 isoform-mediated repression, further providing evidence for a role of HDAC1 in mediating transcriptional repression. The ELM2 domain is also implicated in the MIER1-induced recruitment of HDAC1 as the minimum sequence sufficient for recruitment of HDAC contained the ELM2 domain along with an additional C-terminal sequence that is often conserved in ELM2 containing proteins including MIER1 protein isoforms (Ding *et al.*, 2003).

#### **1.5.3.2.3 Estrogen Receptor (ER) corepressor function of MIER1 $\alpha$ and implications for the LXXLL & the SANT domains**

Recent studies further outlining the function of MIER1 have concentrated on investigating a unique attribute of the MIER1 $\alpha$  protein isoform: the presence of the nuclear receptor interaction LXXLL motif (Heery *et al.*, 1997; Heery *et al.*, 2001;

Savkur & Burris, 2004). This particular motif, whereby “L” signifies the amino acid leucine, and “X” signifies any amino acid, is represented by the amino acids <sup>413</sup>LQMLL<sup>417</sup> in MIER1 $\alpha$  and is located just three amino acids downstream of the 5' start site of the alpha specific C-terminal end (refer to Figure 1.12). In addition to interaction with nuclear receptors, the LXXLL motif is also characteristic of transcriptional regulation as well as nuclear hormone receptor and other TF coregulators (Dobrzycka *et al.*, 2003; Dong *et al.*, 2006; Li *et al.*, 2007; Plevin *et al.*, 2005).

In light of the involvement of the LXXLL motif in the regulation of nuclear receptors, McCarthy *et al.*, 2008 investigated whether MIER1 $\alpha$  is implicated in ER $\alpha$  signalling, and demonstrated that MIER1 $\alpha$  physically interacts with ER $\alpha$  endogenously. Additional studies from our lab have further illustrated that this molecular interaction surprisingly did not involve the LXXLL domain, but instead the C-terminal portion of the SANT domain (Mercer *et al.*, unpublished data). Moreover, this phenomenon was ligand- independent, i.e. ER $\alpha$  & MIER1 $\alpha$  interacted both in the presence and absence of estrogen, the interaction was stronger in the absence of ligand (McCarthy *et al.*, 2008).

Functional implications for the interaction between MIER1 $\alpha$  and ER $\alpha$  were discovered following experiments involving the overexpression of MIER1 $\alpha$  in estrogen-treated T47D breast carcinoma cell lines (McCarthy *et al.*, 2008). The increased presence of this isoform greatly hampered breast carcinoma cell growth in response to estrogen (E2) and revealed MIER1 $\alpha$  as a possible corepressor of ER signalling (McCarthy *et al.*, 2008). Further studies examining MIER1 $\alpha$ 's repressive



effects on ER $\alpha$  investigated how MIER1 $\alpha$  affected the transcriptional regulation of the E2-regulated pS2 [trefoil factor 1 (TFF1)] promoter. Results demonstrated that only the MIER1 $\alpha$  isoform, not the MIER1 $\beta$  isoform, repressed both the ligand-dependent and ligand-independent transcriptional activation at this promoter through molecular interactions with its LXXLL domain (Mercer *et al.*, unpublished data).

Further investigation into the ligand-independent corepressor function of MIER1 $\alpha$  at the pS2 promoter showed, using ChIP assays, that MIER1 $\alpha$  does not affect ER $\alpha$  recruitment to the promoter, nor does its repression involve the recruitment of HDAC (Mercer *et al.*, unpublished data). These ChIP results did, however, show that MIER1 $\alpha$  repressive function of ER-mediated transcription inhibited recruitment of steroid receptor co-activator 3 (SRC-3) and interfered with the dimethylation of histone H3 lysine 9 (H3K9), two processes involving epigenetic modifications that are strongly involved in transcriptional activation at this promoter (Mercer *et al.*, unpublished data).

#### **1.5.3.2.4. Negative Regulation of the Histone Acetyltransferase (HAT) activity of cAMP response element binding protein (CREB) binding protein (CBP) via MIER1 $\beta$ 's ELM2 Domain**

As MIER1 had previously been implicated in recruiting HDAC1 to the site of transcription initiation, Blackmore *et al.*, 2008 investigated whether or not MIER1 was further involved in regulating HAT activity. GST pull-down assays using  $^{35}\text{S}$  labelled flag-tagged CBP constructs, full-length GST-MIER1 $\beta$  fusion proteins, and various deletion mutants of GST-MIER1 $\beta$  fusion proteins showed that MIER1 interacted with CBP, a known HAT. This interaction involved the N-terminal half of the protein,

which includes the ELM2 and acidic activation domains. Moreover, HAT assays on extracts from HEK 293 cells co-transfected with myc-tagged full-length MIER1 $\beta$  and flag-tagged CBP showed MIER1 $\beta$  inhibited CBP (Blackmore *et al.*, 2008).

#### 1.5.4 MIER1 Expression & Subcellular Localization

Studies investigating *mier1* expression patterns have demonstrated that MIER1 $\beta$  is ubiquitously expressed at very low levels in most human tissues (Paterno *et al.*, 1998, Paterno *et al.*, 2002). There are, however, tissues that displayed above average expression levels, including the spleen, adrenal glands, adult & fetal thymus, small intestine, colon, heart, ovary, thyroid glands and testis. Of these tissues, the testis exhibited the highest amount of expression in both reported studies (Paterno *et al.*, 1998, Paterno *et al.*, 2002). Further investigation of MIER1 $\alpha$  isoform-specific transcript expression revealed that it was not as ubiquitously expressed as its  $\beta$  counterpart. However, its expression patterns did support its role as a nuclear receptor coregulator as MIER1 $\alpha$  expression is restricted to endocrine and endocrine-responsive tissues. With respect to the *mier1* splice variants, various tissue-specific expression patterns were noted. For example, the lung and skeletal muscle only produced MIER $\beta$  transcripts harbouring the N3 N-terminal domain. Moreover, the generation of splice variant transcripts harbouring exon 3A across certain tissues was also shown to display tissue-specific expression patterns, as the brain, and skeletal muscle only expressed transcripts harbouring the 3A exon, and the lungs expressed transcripts lacking the 3A exon. These studies demonstrated that *mier1* undergoes differential promoter usage and tissue-specific regulation of transcription.



Analyses regarding the subcellular localization of the *mier1* protein isoforms have confirmed that both the  $\alpha$  and  $\beta$  isoforms reside in the nucleus, an observation fitting with MIER1's role as a potent transcriptional regulator (Mercer *et al.*, unpublished data; Paterno *et al.*, 1997). Originally it was presumed that only the MIER1 $\beta$  protein isoform had the potential to be situated in the nucleus, as it contains the only functional putative nuclear localization signal (NLS) in humans (Ding *et al.* 2003; Post *et al.*, 2001). As delineated in Figure 1.12, this NLS resides approximately 64 amino acids downstream of the 5' starting point of the 102 amino acid MIER1 $\beta$  specific C-terminal end. However, further examination has demonstrated that both forms are found in the nucleus, even though MIER1 $\alpha$  does not contain a NLS. Ding *et al.*, 2003 postulate that this may be due to a co-transport or piggyback mechanism used by the alpha isoform in order to locate to the nucleus, such mechanisms have been described elsewhere for other proteins (Kang *et al.*, 1994; Zacksenhaus *et al.*, 1999).

### **1.5.5 Breast Cancer & MIER1**

Among Canadian women, breast cancer continues to lead in incidence, with 22, 700 new cases expected in 2009 (Canadian Cancer Society 2009 Cancer Statistics, 2009). Even though animal models have given us great insight into the molecular pathways involved in breast cancer development, there are still many remaining gaps in our understanding of the manifestations and progression of this disease (Ip & Asch, 2000; Shackleton *et al.*, 2006; Stingl *et al.*, 2006; Thompson *et al.*, 2008). To date many signalling pathways have been extensively researched with respect to signal transduction in breast cancer, such as: the estrogen receptor, receptor tyrosine kinase



(RTK), and DNA repair pathways. However, knowledge concerning their regulation and interplay during breast cancer is incomplete (Britton *et al.*, 2006; Schiff *et al.*, 2005; Speirs *et al.*, 2007). Furthermore there is a great need for the characterization of the causative factors and potential regulators underlying the invasive progression of breast cancer cells (reviewed in Gonzalez-Angulo *et al.*, 2007; Thompson *et al.*, 2008).

Recent studies focusing on MIER1 $\alpha$  have implicated this *mier1* isoform as having a role in breast cancer progression. Overexpression of MIER1 $\alpha$  in T47D breast carcinoma cells, in conjunction with estrogen treatment, significantly reduced the ability of estrogen to stimulate anchorage independent growth of these cells (McCarthy *et al.*, 2008). These results suggest that MIER1  $\alpha$  may regulate breast carcinoma growth as the ability to proliferate without attachment is a distinct indicator of cell tumourigenicity (Fukazawa *et al.*, 1995; McCarthy *et al.*, 2008). Moreover, the nature of MIER1 $\alpha$ 's corepressor function and interaction with ER $\alpha$ , as described in section 1.5.3.2.3, has strong implications for MIER1's role in regulating ER $\alpha$  signalling, a receptor whose dysregulation has long been established as playing a fundamental role in breast cancer development and tumourigenesis (Chen *et al.*, 2008; Jones *et al.*, 2008; Speirs & Walker, 2007).

These unique characteristics of MIER1 $\alpha$  in association with breast cancer become even more pertinent, though, in the context of invasive disease. Immunohistochemical analysis of normal breast tissues and breast carcinoma tissues of varying stages demonstrated that as breast cancer cells acquire a more invasive phenotype, the subcellular localization of MIER1 $\alpha$  changes dramatically (McCarthy *et*

*al.*, 2008). Instead of residing in the nucleus, whereby 77% of the normal breast samples examined contained nuclear MIER1 $\alpha$ , MIER1 $\alpha$  was mainly situated in the cytoplasm of cells from invasive tissue, with only 4% of invasive ductal carcinoma (IDC) samples staining for nuclear MIER1 $\alpha$  (McCarthy *et al.*, 2008). This striking change in subcellular localization in invasive breast cancer may have ramifications for MIER1 $\alpha$ 's pivotal function as a nuclear transcription factor coregulator. McCarthy *et al.*, 2008 hypothesized that the change in subcellular localization of MIER1 $\alpha$  from the nucleus to the cytoplasm may inhibit MIER1 $\alpha$  from exerting its gene/chromatin repressor function and suppression of ER $\alpha$ , thereby ridding the cell of an inherent security mechanism and subsequently increasing breast carcinoma cell proliferation (McCarthy *et al.* 2008). Moreover, the ability of MIER1 $\alpha$  to regulate ER $\alpha$  in the absence of ligand strongly parallels the function of another gene with implications in breast cancer development, namely BRCA1, which regulates ER $\alpha$ -stimulated and unliganded ER $\alpha$  activity (Gorski *et al.*, 2009; Zheng *et al.*, 2001).

#### **1.5.6 The Regulation of *mier1***

In order to fully understand the role of MIER1 in the transcriptional regulatory network of the cell, it is necessary to characterize the regulatory factors that control its expression and that induce MIER1 isoforms to exert their functional effects. To date several studies have elucidated various molecules that regulate *mier1* transcription, such as: fibroblast growth factor, Sp1, and even *mier1*'s own protein product, MIER1, which functions to inhibit Sp1-induced transcription in the form of a regulatory

feedback loop (Ding *et al.*, 2004; Paterno *et al.*, 1997). Most of the current knowledge about the regulation of *mier1* at the level of transcription has come from the characterization of the MAEP-P2 promoter.

Regulatory mechanisms involved in MLP-P1 promoter- specific transcription have yet to be investigated. As described in section 1.5.4, expression analysis of transcript splice variants demonstrates that *mier1* promoters are under tissue-specific regulation, which further exemplifies that each promoter can be regulated by distinct transcriptional and environmental factors. Moreover, potential transcriptional activation at the MLP-P1 promoter can result in the incorporation of the commonly skipped exon, exon 3A. This phenomenon expands the network of converging factors implicated in *mier1* expressional regulation, as distinct factors can be involved in controlling the subsequent inclusion of this exon.

## **1.6 The Aims and Principle Objectives of this Study**

The full characterization of promoter regions, by virtue of their function in transcriptional regulation, is not only key to understanding tissue-and stage-specific gene expression regulation, but is now increasingly becoming more important to understanding the development of diseases, such as cancer (Heintzman & Ren, 2007). The unique characteristics of the MLP-P1 promoter, and the possible outcomes of MLP-P1- driven transcription, suggest that this promoter may be implicated in the regulation of *mier1* in breast cancer. To this effect, several hypotheses have been made concerning *mier1* regulation at MLP-P1 and breast cancer:



- 1) **If** the putative NES and/ or putative transmembrane domain specified by exon 3A functions to remove proteins from the nucleus, **then** MLP-P1-driven transcription can produce *mier1* protein isoforms that reside in the cytoplasm.
- 2) **If** during breast cancer progression there is a preferential activation of the MLP-P1 promoter over the MAEP-P2 promoter of *mier1*, **then** this can perpetuate the production of an increased number of cytoplasmic / non-nuclear residing *mier1* isoforms which may serve to explain the reported subcellular localization change of MIER1 $\alpha$  from the nucleus to the cytoplasm in invasive breast cancer.

The primary goal of this study was to characterize the MLP-P1 promoter of *mier1*. Determination of the specific sequences required for activity at this promoter facilitated this study's additional aims, which were to investigate *mier1* promoter activity in breast cancer cells, as well as to identify whether or not transcriptional activation at MLP-P1 can affect the subcellular localization of certain *mier1* isoforms. In order to meet these specific aims, this study focused on the following summarized objectives.

**Objective 1-** *Characterize the MLP-P1 promoter, including the analysis and identification of the minimal nucleotide sequence necessary for promoter activity and the nucleotide sequence necessary for maximal promoter activity in breast cancer cells.*

**Objective 2-** *Search for putative human TFBSs residing in the MLP-P1 promoter region in order to identify whether or not this promoter contains any putative TFBS that interact with transcription factors that are dysregulated in breast cancer.*

**Objective 3-** *Characterize MLP-P1 promoter activity in various cell lines and determine if promoter activity is ER status dependent.*

**Objective 4-** *Compare the activity of the two mier1 promoters in various cell lines to determine if there is preferential usage.*

**Objective 5-** *Investigate whether or not the inclusion of exon 3A in mier1 transcripts affects the subcellular localization of MIER1 $\alpha$ .*

## Chapter 2: Material & Methods

### 2.1 Cell Culture

#### 2.1.1 Cell Lines & Cell Maintenance

Four different cell lines were used in this study, as depicted by the following table:

**Table 2.1: Cell lines**

Cell line	Description	Supplier
HEK 293 Cells	Human embryonic kidney cells transformed with Adenovirus 5 DNA	American Tissue Culture Collection (ATCC); ATCC® #: CRL-1573™
Hs578 T Cells	Mammary gland, carcinoma, ER negative	ATCC; ATCC® #: HTB-126™
MCF7 Cells	Mammary gland, adenocarcinoma, ER positive	ATCC; ATCC® #: HTB-22™
MDA MB 231 Cells	Mammary gland, adenocarcinoma, ER negative	ATCC; ATCC® #: HTB-26™
T47D Cells	Mammary gland, ductal carcinoma, ER positive	ATCC; ATCC® #: HTB-133™



HEK 293 cells , MDA MB 231 cells, and Hs578T cells were grown and maintained in Dulbecco's Modified Eagle Medium (DMEM; Invitrogen) supplemented with 10% Serum [75% calf serum (CS; Invitrogen) and 25% fetal bovine serum (FBS; Invitrogen) ], 1% sodium pyruvate (Na-pyruvate; Invitrogen), and 0.5% penicillin/ streptomycin (Pen/Strep; Invitrogen) as per recommendation of the American Type Culture Collection (ATCC). MCF7 cells were cultured in media containing DMEM supplemented with 10% Serum [75% CS and 25% FBS], 1% Na-pyruvate, 0.5% Pen/Strep, and 0.001% of 4mg/ml insulin/ (Invitrogen). T47D cells were grown in RMP1-1640 media (Invitrogen) supplemented with 10% Serum [75% CS and 25% FBS], 2.383 g/L HEPES (Invitrogen), 0.001% of 4mg/ml insulin, and 0.5% Pen/Strep.

All cells were grown in 100mL plates (Corning) containing 10mL of respective media as described above and kept in an incubator at a temperature of 37°C containing 5% CO<sub>2</sub> levels. Cells were allowed to grow until confluency was reached. The optimal confluency (refer to Table 2.2) is different for every cell line and is dictated by their respective growth patterns and cellular structure. Following growth to optimal confluency (usually between 70-90% confluent; refer to Table 2.2), cells were subcultured at various dilutions depending on the cell line (Table 2.2). Briefly, media from each plate was aspirated and then cells were washed carefully with 5ml of 1 x phosphate-buffered saline (PBS; Sigma). The 1 x PBS was then aspirated off and 1.5ml of 1% trypsin/ [ethylenediamine tetraacetic acid (EDTA)] (Invitrogen) in 1 x PBS (Sigma) was added. Cell culture plates were then rocked back and forth until all cells were no longer attached to the plate surface. Cells were then resuspended in the required amount of media that would yield the respective subcultivation ratio and then

added to a new cell culture plate containing enough media to result in a final volume of 10ml.

Sterile technique was strictly employed for all handling and subculturing of cells, which was performed in a laminar flow-hood (E-614, BioKlone). Prior to any cell culture work, the laminar flow hood was exposed to ultraviolet (UV) light for at least 5 minutes and wiped down with 70% ethanol. All equipment, media, and reagents being brought into the laminar flow hood were wiped down with 70% ethanol. Only specifically designated laboratory coats were worn in cell culture and fresh new gloves were always employed.

**Table 2.2: Cell line optimal confluency & subcultivation ratios**

<b>Cell Line</b>	<b>Maximum Confluency</b>	<b>Sub-cultivation Ratios</b>
<b>HEK 293 Cells</b>	80-90%	1:10
<b>Hs578T Cells</b>	70-80%	1:3
<b>MCF 7 Cells</b>	60-70%	1:3
<b>MDA MB 231 Cells</b>	70-80%	1:3
<b>T47D Cells</b>	60-70%	1:2



## 2.2 Vectors & Constructs

### 2.2.1 Experimental Control Plasmid Constructs

Plasmid constructs were an integral part of this study. The following table details the plasmid constructs that were used as experimental control vectors for various employed experiments:

**Table 2.3: Experimental control plasmids**

Plasmid Construct Name & Manufacturer	Plasmid Construct Description
<b>pGL3 empty vector (Promega)</b>	This 4818bp plasmid vector contains <i>luc +</i> gene, which encodes for the firefly luciferase protein. The <i>luc +</i> gene, however, lacks eukaryotic promoter and enhancer sequences and; therefore, expression of the luciferase protein and subsequent luciferase activity depend on the insertion and proper orientation of a functional promoter sequence upstream of <i>luc +</i> . This particular empty vector, with no additional sequence inserted into its multiple cloning site, was used as a negative control for background luciferase activity.
<b>pRSV <math>\beta</math>-Galactosidase (pRSV<math>\beta</math>-gal) (Promega):</b>	This expression plasmid construct contains the bacterial <i>lacZ</i> gene encoding the $\beta$ -galactosidase enzyme inserted into the pRSV vector. It was used as control for monitoring transfection efficiency. The pRSV vector itself contains the rous sarcoma virus (RSV) promoter that drives transcription of the <i>lacZ</i> gene and subsequent production of the $\beta$ -galactosidase ( $\beta$ -gal) enzyme.
<b>hER<math>\alpha</math> pcDNA3</b>	This expression plasmid construct (-kind gift from Dr. Christine Pratt, University of Ottawa) was designed to express the coding sequence of human ER $\alpha$ . In this particular study it was transfected into the ER negative HEK 293 cell line in order to verify whether or not any observed difference in promoter activity results between HEK 293 cells and the MCF7 ER positive breast cancer cell line was not a function of differing ER status (refer to section 3.2.3). The pcDNA mammalian expression vector is designed for high-level, constitutive expression in a variety of mammalian cell lines and includes the cytomegalovirus (CMV) enhancer-promoter for high-level expression as well, as an amp resistance gene.

### **2.2.2 pCS3+MT*mier1*α & pCS3+MT*mier1*α exon3A**

The expression vectors pCS3+MT*mier1*α and pCS3+MT *mier1*α exon 3A were used in order to investigate the effects of encoding exon 3A and the subsequent incorporation of the putative NES on the subcellular localization of *mier1* protein isoforms. Full length *mier1*α-specific isoform genomic sequence stemming from the MLP-P1 promoter (GeneID: 57708; accession number NM\_001077703) was engineered and cloned into the pCS3+MT empty vector Z. Ding in our laboratory (Ding, 2004). The pCS3+MT empty vector is a myc-tagged mammalian expression vector that was kindly provided by Dr. David Turner, University of Michigan, whereby the myc tag constitutes six N-terminal repeats of the amino acid residues MEQKLISEEDLNE from the c-MYC protein.

Furthermore, the pCS3+MT *mier1* α exon 3A (GeneID: 57708; accession number NM\_001077702) was constructed by Corinne Mercer in our laboratory and includes the full-length *mier1*α-specific isoform genomic sequence stemming from the MLP-P1 promoter and containing the encoding sequence of exon 3A.

### **2.2.3 *mier1* Luciferase Reporter Gene Deletion Constructs**

Luciferase reporter gene constructs containing specific sequences of *mier1* promoter regions inserted into the pGL3 empty vector were used to analyze and compare *mier1* promoter activity. All constructs, except for MLP-P1 (-91) pGL3 and MLP-P1 (-44) pGL3, were cloned and engineered by Zhihu Ding, PhD (Ding, 2004). Briefly, this included the sub-cloning of *mier1* promoter specific cDNA, which was

originally isolated from human primary ectocervical cells, into the pCR2.1 vector. The isolated cDNA included 4077 bp of *mier1* genomic sequence cDNA that spanned the MLP-P1 promoter region and 3045 bp of *mier1* genomic sequence cDNA spanning the MAEP-P2 promoter region. Further deletion constructs were then generated through PCR- amplification of specific regions of each promoter using the primer pairs listed in Tables 2.4 & 2.5, and using the previously mentioned constructed plasmids as templates. These amplification products were then each cloned into separate pCR2.1 vectors. Each pCR2.1 vector was then digested with distinct restriction endonucleases in order to facilitate the directional cloning and ligation of the cloned sequences into the pGL3 empty vector to produce the final *mier1*-promoter specific luciferase reporter gene deletion construct (Ding, 2004). For further in-depth details on full cloning procedures performed in our laboratory, refer to either section 2.4 or Ding, 2004.



**Table 2.4: Primers used for cloning & engineering of MLP-P1 luciferase reporter gene deletion constructs**

<b>MLP-P1 Luciferase Reporter Gene Construct</b>	<b>5' Primer Sequence</b>	<b>3' Primer Sequence</b>
<b>MLP-P1(-1708) pGL3</b>	5'TGCAGGTTGGT AGCCTAGAAGCAACA 3'	5'TCCGTCTTGTC TGCATTGAACC 3'
<b>MLP-P1 (-1077) pGL3</b>	5'GCTGTGTGCTT TTCTACAGTCTTGTTTC 3'	5'TCCGTCTTGTC TGCATTGAACC 3'
<b>MLP-P1(-742) pGL3</b>	5'CTCGAGTGCAA CGGCACGATCT 3'	5'TCCGTCTTGTC TGCATTGAACC 3'
<b>MLP-P1(-185) pGL3</b>	5'CCCAAGCTCCT CCGTTAGCG 3'	5'TCCGTCTTGTC TGCATTGAACC 3'
<b>MLP-P1(-91) pGL3</b>	5'CGGTTGGCTGG CGGCACG 3'	5'TCCGTCTTGTC TGCATTGAACC 3'
<b>MLP-P1(-44) pGL3</b>	5'GAGGGCAGAGG GTTGGTGGAG 3'	5'TCCGTCTTGTC TGCATTGAACC 3'
<b>MLP-P1(+37) pGL3</b>	5'CTCCCAGGCTCT GAGTCTCC 3'	5'TCCGTCTTGTC TGCATTGAACC 3'

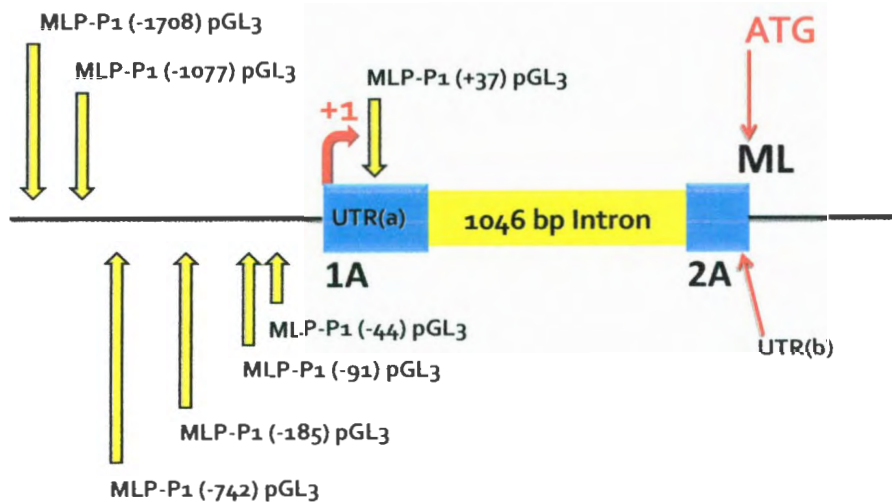
**Table 2.5: Primers used for cloning & engineering of MAEP-P2 luciferase reporter gene deletion constructs**

<b>MAEP-P2 Luciferase Reporter Gene Construct</b>	<b>5' Primer Sequence</b>	<b>3' Primer Sequence</b>
<b>MAEP-P2 (-1316) pGL3</b>	5'GACTGTCTGTAG ACTCTTTTCC 3'	5'CGTACTGCCGGGT CACATCTCC 3'
<b>MAEP-P2 (-312) pGL3</b>	5'ACGTATTTTTCC TCTGCTGTGTCA 3'	5'CGTACTGCCGGGT CACATCTCC 3'
<b>MAEP-P2 (-68) pGL3</b>	5'TTTCCCTCCAGT CCAGCCCAGCCG 3'	5'CGTACTGCCGGGT CACATCTCC 3'
<b>MAEP-P2 (+28) pGL3</b>	5'AGTGGCGGCGG GAGCGGCAGAGA 3'	5'CGTACTGCCGGGT CACATCTCC 3'

### **2.2.3.1 MLP-P1 Luciferase Reporter Gene Deletion Constructs**

Seven distinct MLP-P1 promoter-specific luciferase reporter gene deletion constructs were used throughout the duration of this study. Each plasmid construct includes, inserted into the multiple cloning site of the pGL3 empty vector, the entire DNA sequence stemming from the respective 5' starting nucleotide (as dictated by the 5' primer used for its construction, refer to Table 2.4) up until seven base pairs upstream of the ATG start site, as dictated by the 3' primer used to engineer each construct (refer to Table 2.4 for the exact sequences of the primers used to engineer these reporter gene constructs). It is also important to note that each MLP-P1 luciferase reporter gene deletion construct includes the 1046bp intron in the MLP-P1 promoter region, the exact location of which is illustrated in Figures 2.1 & 2.2 (refer to Appendix 2 for the exact nucleotide of this specific intron). The nucleotide counts represented in each luciferase reporter gene deletion construct name are relative to the putative TSS of MLP-P1. As this promoter has yet to be completely analyzed, the actual TSS remains to be fully delineated. The putative TSS resides at the the first nucleotide incorporated into the MLP-P1 5' UTR, and thus the first nucleotide of exon 1A. Moreover, Figure 2.1 represents a schematic of the relative location of 5' starting position of the insertion sequences incorporated into each MLP-P1 luciferase reporter gene deletion construct. The full nucleotide sequences of the construct inserts with reference to the overall MLP-P1 promoter region nucleotide sequence are depicted in Figure 2.2 for four of the plasmid constructs containing sequences of the proximal promoter area. Refer to Appendix 1 for the exact insertion sequences and 5' start positions of the three additional constructs not included in Figure 2.2 (-MLP-P1(-1708) pGL3, MLP-P1(-1077) pGL3, and the MLP-P1 (-742) pGL3 constructs).





**Figure 2.1: Schematic of *mirl* MLP-P1 promoter region with reference to the sequences incorporated into the MLP-P1 luciferase reporter gene deletion constructs**

Schematic depicts the MLP-P1 promoter region, as well as the two exons (represented by the blue boxes) of *mirl* that are incorporated into *mirl* transcripts following transcriptional activation at the MLP-P1 promoter. Exon 1A or UTR (a) is 137bp long, while exon 2A, which encodes for the untranslated region UTR(b) of MLP-P1 specific mRNA transcripts, is 101bp long. This exon also encodes for the first two amino acids (methionine and leucine) of the protein. The putative transcription start site, or +1, is situated at the commencement of exon 1A and is 1278 bp upstream of the ATG translation start site, if the size of the intron (represented by the yellow box) is included. Without the intron included in the count, the TSS is 232bp upstream of the ATG translation start site. The 5' start positions of the nucleotide sequences contained in each respective luciferase reporter gene deletion construct are represented with the yellow arrows.

```

(-249/-1528) gtggcgacca gctggggagt ggtgcaccac cccttttttt ggccgcctct gaagtccttg

MLP-P1(-185) pGL3
(-189/-1468) taccctcaag ctctccggtt agcggctcgg gccgaggctc cggaatgttt gccgggcgctc

MLP-P1(-91) pGL3
(-129/-1408) atggcgacgg tggagccctg gctcaacaag cggccgcgctg gttggctggc ggcacgaggc

MLP-P1(-44) pGL3
(-69/-1348) cgaggaggag ggcggaggcg gaggggaggg cagagggttg gttggagctgg aggaagctcc

Putative TSS / Exon 1A
(-9/-1288) ggacgacgaC TGAAGAAGG AGGCGGGCGG CCCGGGCCTC AGGCCCCCTCC CAGGCTCTGA

MLP-P1(+37) pGL3
(+52/-1228) GTCTCCCGGC TGCAGGCGGA TGGATGGGGC TTCTTCAGGC GGTGGCGGCA GCAGCGAAGG

(+112/-1168) TGGCGGCGGC AGCAGCGGCA GCGGCT.....1046bp Intron.....

Exon 2A
(+1181/-95) ...ATGGTGT GGTGCTCGA TTCTCCAGT GCCTGGCTGA GTTTCGGACG TGGTTAAGAA

3'end of reporter gene inserts
(+1241/-38) CCAACTGGTT GAGGTTCAAT GCAGACAAGA CGGATGTGAT GCTG

```

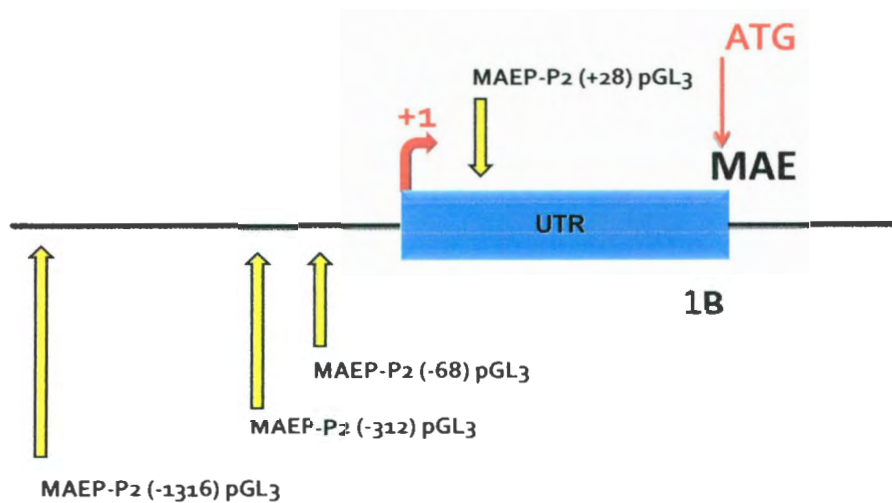
**Figure 2.2: DNA sequence of the MLP-P1 promoter proximal region**

The above figure depicts the *mier1* nucleotide sequence (Entrez Gene ID: 57708) of the MLP-P1 promoter proximal region spanning 1528bp upstream of the ATG translational start site and 249 bp upstream of the putative TSS. The first number on the left side is the nucleotide count with respect to the putative TSS of MLP-P1 where as the second number is the count with respect to the ATG translation start site. The sequence of the 1046bp intron is not included (Refer to Appendix 2 for the complete DNA sequence of the intron). Four out of the seven of the MLP-P1 specific reporter construct 5' sequence start sites are indicated. (Refer to Appendix 1 for the location of the other three constructs). The putative TSS, as well as the start of exons 1A and 2A, are denoted along with the ATG translational start site, which is typed in red. Uppercase letters demarcate the nucleotides in either exon 1A or 2A. The location whereby all MLP-P1 luciferase reporter gene deletion construct insertion sequences terminate at the 3'end is also denoted, and resides 7 bp upstream of the MLP-P1 ATG translation start site.

### 2.2.3.2 MAEP-P2 Luciferase Reporter Gene Deletion Constructs

There were four different MAEP-P2 luciferase reporter gene deletion constructs used in this study. These constructs include, in the insertional cloning area of pGL3, all of the DNA sequence stemming from the respective 5' starting position (as dictated by the 5' primers used, refer to Table 2.5) to six nucleotides upstream of the ATG start site, as dictated by the 3' primer used in the engineering of each construct (Table 2.5). Figure 2.3 is a schematic of the MAEP-P2 luciferase reporter gene deletion constructs relative to the MAEP-P1 promoter region. The numbers incorporated into the name of each MAEP-P2 luciferase reporter gene deletion construct represent the count of the 5' nucleotide relative to the TSS of MAEP-P2. As this promoter had been already characterized (Ding, 2004; Ding *et al.*, 2004), the actual known TSS is denoted (Fig 2.3 & 2.4). The actual nucleotide sequences inserted into the plasmid constructs with reference to the overall MAEP-P2 promoter region are represented in Figure 2.4 for three out of the four employed MAEP-P2 luciferase reporter gene deletion constructs. Refer to Appendix 3 for the full nucleotide sequence of the insert cloned into the largest MAEP-P1 luciferase reporter gene deletion construct [MAEP-P2 (-1316) pGL3] with reference to the overall MAEP-P2 promoter region.





**Figure 2.3: Schematic of the *mier1* MAEP-P2 promoter region with reference to the sequences incorporated into the MAEP-P2 luciferase reporter gene deletion constructs**

Diagram depicts the MAEP-P2 promoter region, as well as the exon of *mier1* (exon 1B) that results following transcription at this specific promoter. Exon 1B, represented by the blue box, is 121 bp long and encodes the UTR as well as the first three amino acids of the protein (methionine, alanine, and glutamic acid). The transcription start site, represented by +1, begins 148 bp upstream of the ATG translation start site. The 5' start positions of the nucleotide sequences contained in each respective luciferase reporter gene deletion construct are represented with the yellow arrows.

(-416/-565)	caaggcataa	ccgctttcaa	agcacaqffc	ttgcgcgcqt	tgcatacacag	caacatccgt	
(-356/-505)	ttttcagcaa	atgcatttca	aaaacgacct	acatgtaaaa	tattacgtat	ttttctctg	<b>MAEP-P2 (-312) pGL3</b>
(-296/-445)	ctgtgtcaat	gctgqatgtg	actccitttg	cactgttctt	tgacctctct	gatacgttga	
(-236/-385)	caaccccttc	cgectqaggt	ggcgqatcag	ctggagccag	cgaagcgccc	cgcgcggttg	
(-176/-325)	cccacctcct	cccacaccca	ccttqaactc	ggcccccctc	ctcttcccgg	qgaqgqctgg	
(-116/-265)	ccgcgqggcc	gcgcgcgcgc	ccctgcctcg	gcgcgtgctc	gctggtcttt	tcctccagtc	<b>MAEP-P2 (-68) pGL3</b>
(-56/-205)	ccagcccagc	cggggcgcgc	cgagggggcg	gagtggggtg	tggtaggcgc	gctcggtcgg	<b>+1</b>
(+5/-145)	CTCCTGCGCG	TTCCCGCCGA	GGCAGTGGCG	GCGGGAGCGG	CAGAGACGGC	AGCGGCCCGA	<b>MAEP-P2 (+28) pGL3</b>
(+65/-85)	GTCCCGTTGC	TGAGTCTCAC	ATCCGGGTTC	TGGCUGTGAC	CCAGCTGCGG	CCGCCGCCGA	
	<b>3 end of reporter gene inserts</b>						
(+125/-25)	GATGTGACCC	GGCAGTACGG	CAAATATGCG	G			

**Figure 2.4: DNA sequence of the MAEP-P2 promoter proximal region**

The above figure depicts the *mier1* nucleotide sequence (Entrez Gene ID: 57708) of the MAEP-P2 promoter proximal region spanning 565bp upstream of the ATG translational start site and 416bp upstream of the TSS. The first number on the left side is the nucleotide count with respect to the TSS of MAEP-P2, where as the second number is the count with respect to the ATG start site. Three out of the four MAEP-P2 specific reporter construct 5' sequence start sites are indicated. (Refer to Appendix 3 for the location of the additional upstream luciferase reporter deletion construct MAEP-P2 (-1316) pGL3 location.) The TSS/ beginning of exon 1b is indicated, along with the ATG translational start site, which is denoted in red. Uppercase letters demarcate the nucleotides in exon 1b. The location whereby all MAEP-P2 luciferase reporter gene deletion construct insertion sequences terminate at the 3' end is also denoted, and resides 6 bp upstream of the MAEP-P2 ATG translation start site.

### 2.3 Plasmid Construct Preparation, Purification, Quantification, & Storage

In order to generate or replenish laboratory stocks of the various reporter gene deletion constructs, XL Blue chemically competent *Escherichia coli* (*E. coli*) cells (XL blue cells) (Stratagene Inc.) containing the respective reporter gene deletion construct were streaked on a Luria Broth (LB) ampicillin (amp; Sigma-Aldrich) plates [5g peptone (MP Biomedical), 2.5g yeast extract (Fisher Scientific), 5g NaCl (Fisher Scientific), 7.5g Agar (Difco Laboratories), 500ml dH<sub>2</sub>O, 50µg/ ml amp, autoclaved ]. XL Blue cells containing respective reporter gene deletion constructs were obtained from either transforming new XL Blue cells with the reporter plasmids or collecting bacteria from glycerol stocks [20% glycerol (Fisher Scientific), and 80% LB media (5g peptone , 2.5g yeast extract , 5g NaCl, 500ml dH<sub>2</sub>O, autoclaved)] harbouring bacterial cells previously transfected with the respective reporter plasmid. Briefly, XL Blue cell transformation involved incubating 5ng of reporter construct DNA with 100µl of XL Blue cells for 30 minutes on ice. Cells were then heat-shocked at a temperature of 42°C for 45 sec and then put on ice for another 2 minutes. Two hundred and fifty microlitres of LB media was then added to the cells and shaken for 1 hour at 250 rpm at 37°C. One hundred microlitres of this LB/ cell mixture was then plated on a LB amp plate and incubated overnight at 37°C. In the case of collecting bacteria from glycerol stalks, approximately 1 drop of glycerol stock sample was streaked onto a LB amp plate and incubated overnight at 37°C. The reason why LB amp plates were used was because only bacteria harbouring the plasmid constructs, which encode an amp-resistance gene, would be able to grow in the presence of amp.



Following the overnight growth of the bacterial colonies, one or two colonies were inoculated into 150ml of LB media containing 750 $\mu$ l of amp and were left to incubate over night at 37°C while horizontally shaking at 250rpm. The following day, the constructs were collected and purified using a Nucleobond® PC500 EF plasmid maxi prep kit (Catalogue # 740 550; Clontech Laboratories Inc.) as per manufacturer's protocol. Maxi-prep kit purification expected yield is usually between 100-500 $\mu$ g of DNA that is then subsequently resuspended in TE-EF buffer supplied by the Nucleobond Maxi-Prep Kit. Reporter gene deletion plasmid constructs were then quantified using Quant-iT™ dsDNA BR Assay Kit (Q-Bit) (Invitrogen, Inc) following manufacturer's protocols. Following quantification, all reporter gene deletion construct stocks were either diluted or re-precipitated in order to be kept at a 1 $\mu$ g/ $\mu$ l DNA concentration and were then stored at 4°C. Furthermore, verification of the integrity of each plasmid construct was performed by DNA gel electrophoresis (refer to section 2.4.2) whereby both 80ng and 160ng of each plasmid was run on a 0.8% agarose gel and analyzed.

#### **2.4 Reporter Gene Deletion Construct Design, Cloning, & Construction**

For this study, two additional deletion constructs were engineered in order to further investigate the essential nucleotide sequences involved in MLP-P1 promoter activity. These extra reporter constructs were MLP-P1(-91) pGL3 and MLP-P1(-44) pGL3, which were illustrated in section 2.2.3. Figure 2.5 illustrates the steps involved in the design and generation of these additional luciferase reporter gene deletion constructs.



**Figure 2.5: Flow-chart for reporter gene deletion construct design, cloning, & construction**

Engineering of the MLP-P1 luciferase reporter gene deletion constructs MLP-P1 (-91) pGL3 and MLP-P1 (-44) pGL3 followed the steps depicted in this figure.

### 2.4.1 *in silico* Promoter Analysis & Primer Design

Four kilobase pairs of sequence upstream of the ATG translation start site, 2722bp upstream of the putative TSS (-not including the 1046bp intron) of the MLP-P1 promoter was subjected to *in silico* promoter analysis using the following programs:

Algorithm Program	Website
Transcription Element Search System	<a href="http://www.cbil.upenn.edu/cgi-bin/tess/tess">http://www.cbil.upenn.edu/cgi-bin/tess/tess</a>
Transcription Factor Binding Site (85% and 80% homology searches)	<a href="http://mbs.cbrc.jp/research/db/TFSEARCH.html">http://mbs.cbrc.jp/research/db/TFSEARCH.html</a>
Gene Regulation-Ali Baba 2.1	<a href="http://www.gene-regulation.com/">http://www.gene-regulation.com/</a>
CONSITE	<a href="http://asp.ii.uib.no:8090/cgi-bin/CONSITE/consite">http://asp.ii.uib.no:8090/cgi-bin/CONSITE/consite</a>

Following *in silico* analysis, and using previously collected preliminary results as a reference (refer to Appendix 4 for preliminary results), 5' primers and a 3' primer were designed to anneal to specific regions of the MLP-P1 promoter DNA sequence between the MLP-P1 (-185) pGL3 and the MLP-P1(+37) pGL3 sequences. Table 2.6 lists the exact sequences of the primers developed. To facilitate eventual directional cloning into the pGL3 empty vector, these primers were further engineered with a Bgl II restriction enzyme recognition site on the 5' primer and a HindIII restriction enzyme site on the 3' primer (refer to Table 2.8). Furthermore, the newly designed primers also contained several 5' flanking nucleotides (either cytosine (C), guanosine (G), thymine (T), or adenine (A)) to enable the successful cleavage of the cloned product by the restriction enzymes as described by *New England BioLabs* (*New England BioLabs*



2009-10 Catalog & Technical Reference) with respect to the kinetic and enzymatic properties inherent to each respective restriction enzyme (Table 2.6).

**Table 2.6: Primers designed for the construction of MLP-P1 (-91) pGL3 and MLP-P1 (-44) pGL3 luciferase reporter gene deletion constructs**

<b>MLP-P1 Luciferase Reporter Gene Deletion Construct</b>	<b>5' Primer Sequence</b>	<b>3' Primer Sequence</b>
<b>MLP-P1 (-91) pGL3</b>	5' <i>CTT</i> <u>AGATCT</u> CGGTT GGCTGGCGGCACG 3'	5' <i>CTA</i> <u>AGCTT</u> GTCTT GTCTGCATTGAACC 3'
<b>MLP-P1 (-44) pGL3</b>	5' <i>AAG</i> <u>AGATCT</u> GAGGGC AGAGGGTTGGTGGAG 3'	5' <i>CTA</i> <u>AGCTT</u> GTCTT GTCTGCATTGAACC 3'

Underlined sequences denote the restriction enzyme recognition sites (AGATCT = Bgl II restriction enzyme site; AAGCTT = HindIII restriction enzyme site). The italicized letters represent the additional nucleotides incorporated to enable successful cleavage by the restriction enzymes.

#### 2.4.2 Polymerase Chain Reaction

Polymerase chain reaction (PCR) was performed using Platinum<sup>®</sup> *Taq* DNA Polymerase High Fidelity (Invitrogen) according to the manufacturer's protocol in order to produce the *mier1* MLP-P1 reporter gene deletion construct-specific insertion sequence. This particular enzyme, designed for cloning purposes, was chosen based on its ability to ensure the accuracy of the original DNA sequence template (Invitrogen). The template for the PCR reactions was the previously constructed pCR2.1 vector containing the originally isolated 4077bp cDNA insert from human primary ectocervical cells (refer to section 2.2.3). Primers used were the MLP-P1(-91) pGL3 - specific and the MLP-P1 (-44) pGL3-specific primers described in section 2.4.1. Briefly, 48µl of master mix (refer to Table 2.7 for components of Platinum<sup>®</sup> *Taq* DNA Polymerase High Fidelity Master mix) was added to 2µl of 1ng/µl template plasmid construct in a 0.5ml thin-wall PCR tube (Fisher Scientific). For negative controls, 2µl of dH<sub>2</sub>O was used instead of the template plasmid construct. Contents were gently mixed, centrifuged, and then placed in a thermal cycler (Eppendorf Authorized Thermal Cycler) and subjected to the PCR program outlined in Table 2.8.



**Table 2.7: Platinum<sup>®</sup> Taq DNA Polymerase High Fidelity master mix reagents**

Reagent	Amount/tube (μl)	Amount/Master Mix (x 4.5) (ie: for 4 samples)
<b>10x High Fidelity PCR Buffer (Invitrogen)</b>	5	22.5
<b>50mM MgSO<sub>4</sub> (Invitrogen)</b>	2	9
<b>10mM dNTPs (Invitrogen)</b>	1	4.5
<b>100μg/μl of 5' primer</b>	1.25	5.6
<b>100μg/μl of 3' primer</b>	1.25	5.6
<b>Sterile dH<sub>2</sub>O</b>	37.5	168.8
<b>Platinum<sup>®</sup> Taq High Fidelity</b>	0.2	0.9

Master Mixes were calculated using 4.5 in the hypothetical 4 sample master mix so as to account for pipetting error. Platinum High Fidelity Taq DNA Polymerase was added last to the master mixes, following mixture of the other reagents.

**Table 2.8: Thermocycler parameters for PCR using Platinum<sup>®</sup> *Taq* DNA Polymerase High Fidelity**

Temperature (°C)	Time (sec)	Number of Cycles
94	120	1
55	30	
68	33	30
94	30	
55	30	
68	420	1
30	1	
4	HOLD	

The 94°C temperature incubation was for denaturing of the template, the 55°C temperature incubation ensured annealing of primers, while the 68°C temperature incubation allowed for extension.

Following the initial PCR reaction using Platinum® *Taq* DNA Polymerase High Fidelity, 0.2µl of normal *Taq* DNA Polymerase (Invitrogen) was added to each sample thereby enabling an addition of a single deoxyadenosine (A) residue to the 3' ends of each cDNA product. This A-overhang is necessary for the TOPO TA cloning step described in the next section. The PCR products were incubated with *Taq* DNA Polymerase for 10 minutes at 72°C. PCR product sizes were verified by DNA gel electrophoresis of 5µl of the PCR product diluted in 5µl of sterile dH<sub>2</sub>O and 2µl of '1 x blue juice'[Diluted from 10 x blue juice: 0.25% bromophenol blue (BioRad), 0.25% xylene cyanol FF (BioRad), 25% Ficoll (Amersham GE Healthcare)]. All gels were made up of 1% agarose (Invitrogen) and 1 x TBE buffer [-diluted from 10 x TBE {108g Tris base (Fisher Scientific), 55g boric acid (Fisher Scientific), and 40ml 0.5M EDTA pH 8.0 (Sigma) per litre of 10 x TBE, autoclaved}]. Gels were run at a voltage of 115V for 40 minutes.

#### **2.4.3 TOPO Cloning of *mier1* MLP-P1 Promoter Specific Insertion Sequences into the pCR2.1 Vector**

Following verification of the PCR amplicon integrity and size using DNA gel electrophoresis, these products were then cloned into the pCR2.1 vector/ pCR®2.1-TOPO (Invitrogen) using a TOPO TA Cloning® Kit (Invitrogen) as per manufacturer protocol. The pCR2.1 vector provided in this kit, pCR®2.1-TOPO, is designed specifically to contain overhanging 3' deoxythymidine (T) residues which are able to efficiently ligate with the (A) overhangs produced following addition of *Taq* Polymerase to the PCR-amplified inserts described in section 2.4.2.



Briefly, 2µl of freshly generated PCR product containing the MLP-P1 promoter reporter gene deletion construct specific insertion sequence was each respectively added to a solution containing 2µl of salt solution [1.2M NaCl, 0.06M MgCl<sub>2</sub>], 7µl of sterile dH<sub>2</sub>O, and 1µl of the pCR<sup>®</sup>2.1-TOPO vector (-all of which was provided by the TOPO TA Cloning<sup>®</sup> Kit (Invitrogen)) . The reaction was mixed gently and then left to incubate for 5 minutes at room temperature. Following this incubation period, the reaction tubes were placed on ice and 2µl of each respective TOPO<sup>®</sup> Cloning reaction was added to a vial of chemically competent One Shot<sup>®</sup> INV *E. coli* and mixed gently. This reaction was then left to incubate on ice for 30 minutes and then heat-shocked for 30 seconds at 42°C. Immediately after the heat-shock step, the reaction tubes were placed on ice and then 250µl of room temperature SOC medium [2% tryptone, 0.5% yeast extract, 10mM NaCl, 2.5mM KCl, 10mM MgCl<sub>2</sub>, 10mM MgSO<sub>4</sub>, 20mM glucose] (Invitrogen; catalogue # 15544-034) was subsequently added to each respective tube. The tubes were then horizontally shaken at 200 rpm for 1 hour at 37°C. Following this incubation, both 50µl of transformation alone and 10µl of transformation with 20µl of SOC media (used to ensure even spread of bacterial cells) for each respective transformation was spread on a pre-warmed LB amp plate and let to incubate overnight at 37°C.

#### **2.4.4 Colony Screening, Mini-Preparation & Purification, DNA Quantification, and Sequencing of MLP-P1 (-91) pCR2.1 and MLP-P1 (-44) pCR2.1**

Positive colonies containing the pCR<sup>®</sup>2.1-TOPO vector ligated with MLP-P1 promoter region DNA sequence were screened using conventional PCR with *Taq* DNA Polymerase (Invitrogen). The PCR 5' primer that was used was a primer designed specifically to anneal the vector exactly 39bp upstream of the insertion site

(Sequence: 5'GATCCACTAGTAACGGCCGCC 3'), where as the PCR 3' primer used was designed to anneal to the 3' end of both the MLP-P1 reporter gene deletion construct specific insertion sequences (Sequence: 5'CTAAGCTTGTCTTGTCTGCATTGAACC 3') . Briefly, PCR was performed using 25µl reactions whereby, while employing aseptic technique, the bacterial colony to be screened was touched by a pipet tip which was then submersed into an already prepared 25µl PCR reaction (refer to Table 2.9 for description of PCR Master Mix reagents). This same pipet tip was then used to gently streak a pre-warmed LB amp plate in order to grow up and store the possible positive colony. The streaked LB amp plates were placed in 37°C temperature and left to grow over night and the PCR reaction tubes were then placed in a thermal cycler and subjected to the PCR program outlined in Table 2.10. PCR amplicons were then analysed by DNA gel electrophoresis.

**Table 2.9: *Taq* DNA Polymerase master mix reagents for colony screening**

Reagent	Amount/tube (μl)	Amount/Master Mix (x 4.5) (ie: for 4 colonies)
<b>10x PCR Buffer (Invitrogen)</b>	2.5	22.5
<b>50mM MgCl<sub>2</sub> (Invitrogen)</b>	0.75	9
<b>10mM dNTPs (Invitrogen)</b>	2	4.5
<b>100μg/μl of 5' primer</b>	1	5.6
<b>100μg/μl of 3' primer</b>	1	5.6
<b>Sterile dH<sub>2</sub>O</b>	17.65	168.8
<b><i>Taq</i> DNA Polymerase</b>	0.1	0.9

Master Mixes were calculated using 4.5 in the hypothetical 4 colony master mix so as to account for pipetting error. *Taq* DNA Polymerase was added last to the master mixes, following mixture of the other reagents. Then, the pipet tip that came into contact with the bacterial colony to be screened was submerged in a tube containing 25μl of the PCR master mix.



**Table 2.10: Thermocycler parameters for PCR colony screening using *Taq* DNA Polymerase**

Temperature (°C)	Time (sec)	Number of Cycles
94	120	1
55	45	30
72	45	
94	45	
55	30	1
72	420	
30	1	
25	HOLD	

The 94°C temperature incubation was for denaturing of the template, the 55°C temperature incubation ensured annealing of primers, while the 72°C temperature incubation allowed for extension.

Bacterial colonies that resulted in the expected PCR product following colony screening were selected to further prepare and purify. Two positive bacterial colonies for each MLP-P1 reporter gene deletion construct were inoculated into 5ml of LB media containing 25µl of amp and left to grow over night at 37°C while horizontally shaking at 250 rpm. The following day, the MLP-P1 reporter gene deletion constructs in the pCR2.1 vector (-denoted separately as: MLP-P1 (-91) pCR2.1 and MLP-P1 (-44) pCR2.1) were purified using the Wizard® *Plus* SV Minipreps DNA Purification System (Promega; catalogue # A1460) following the manufacturer's protocol. Each construct was then quantified using the Q-Bit as per manufacturer's protocols. Verification of the integrity of each plasmid construct was performed by DNA gel electrophoresis whereby both 80ng and 160ng of each plasmid was run on a 0.8% agarose gel and analyzed. Following DNA construct quantification, the insertion sequences of MLP-P1 (-91) pCR2.1 and MLP-P1 (-44) pCR2.1 were completely sequenced in order to verify the exact nucleotide composition of the newly generated constructs. All sequencing was performed at Memorial University's Genomic and Proteomics (GAP) facility and using an Applied Biosystems 3730 (ABI 3730) automated 48-capillary DNA analyzer (ABI 3730) (Applied Biosystems).

#### **2.4.5 Restriction Digest, Gel excision, Ligation, and Directional Cloning of MLP-P1 (-91) pCR2.1 & MLP-P1 (-44) pCR2.1 inserts into the pGL3 empty vector**

Once MLP-P1 (-91) pCR2.1 and MLP-P1 (-44) pCR2.1 were confirmed as having the proper insertion sequence, they were digested with Bgl II (Invitrogen) and Hind III (Invitrogen) restriction enzymes in order to isolate the insertion sequences in

these constructs. The pGL3 empty vector (the ligating vector), was also digested with the same enzymes. The total amount of DNA digested was 5µg for the pGL3 empty vector and 12µg for each respective pCR2.1 MLP-P1 construct. Digestions were carried out in a total volume of 40µl, including: vector DNA, 3 x excess of each restriction enzyme, and 10% of 10 x React 2 Buffer (Invitrogen). Reaction 2 buffer was chosen as both enzymes were able to function optimally in this buffer. Each reaction was mixed and left to incubate at 37°C for 1 hour. The digested products for both the pGL3 empty and MLP-P1 pCR2.1 vectors were loaded on a 1% agarose /1 X TBE gel for 40 min at 115 volts. The gel was then visualized under low UV light and the area containing the desired DNA bands was excised from the gel. The digested inserts and digested pGL3 empty vector were then extracted and purified using a QIAquick Gel Extraction Kit (Qiagen, catalogue # 28704) as per manufacturer's instructions. Following purification, the purified Bgl II and Hind III digested inserts from MLP-P1 (-91) pCR2.1 and MLP-P1 (-44) pCR2.1 and the purified Bgl II and Hind III digested pGL3 empty vector were quantified by the Q-Bit assay kit. Once the concentration of each was known, then these digested products were ligated using T4 DNA Ligase (Invitrogen, Inc). Briefly, a molar ratio of 3:1 of Bgl II and Hind III digested insertion sequence: Bgl II and Hind III digested pGL3 empty vector (approx. 50ng of total DNA) was added to 4µl of 5 x T4 ligase buffer (Invitrogen), 1 µl of T4 DNA ligase, and enough dH<sub>2</sub>O to yield a final reaction volume of 20µl. This reaction was incubated overnight at room temperature. The following day, ligation products were then transformed into XL Blue *E. coli*. using the same method as described in section 2.3.



#### **2.4.6 Colony Screening, Mini-Preparation & Purification, DNA Quantification, and Sequencing of MLP-P1 (-91) pGL3 and MLP-P1 (-44) pGL3**

Bacterial colonies were then screened in order to find the specific colonies that picked up the pGL3 vectors harbouring the MLP-P1 promoter sequence inserts using conventional PCR following the same procedure as described in section 2.4.4. The 5' primer used to screen these colonies was designed to anneal exactly 62 bp upstream of the multiple cloning site of pGL3 empty vector (sequence: 5'CTAGCAAAATAGGCTGTCCC 3'), while the 3' primer used was the same as described in section 2.4.4. Once positive colonies were found the specific colonies were inoculated in LB media and collected and purified using the Wizard® *Plus* SV Minipreps DNA Purification System as described in section 2.4.4. The subsequent DNA quantification and sequencing were also performed as described earlier. Once sequencing confirmed that the final pGL3 constructs contained the desired sequence, these constructs, denoted: MLP-P1 (-91) pGL3 and MLP-P1(-44) pGL3 respectively, were then further prepared, purified, quantified, and stored as described in section 2.3.

#### **2.5 Transient Transfection**

Transfection protocols were dependent on the cell line used, as conditions for optimal growth and transfection efficiency varied accordingly.

### **2.5.1 HEK 293, Hs578T, MDA MB 231, and T47D Cell Transfection**

Approximately 18 hours prior to transfection, either HEK 293 cells, Hs578T, MDA MB 231 cells, or T47D cells were seeded at various densities (refer to Table 2.11) in 6-well plates (Corning) containing 2ml of either HEK 293 cell supplemented media, MDA MB 231 supplemented media, T47D supplemented media, or Hs579T supplemented media per well. Directly preceding transfection, the supplemented media was replaced with 1.5ml of fresh serum free media (SFM) depending on the cell line being used, ie: HEK 293, MDA MB 231, and Hs578T SFM contained DMEM, and 1% Na- pyruvate, while T47D SFM contained RPMI, 1% HEPES, and 0.001% of 4mg/ml insulin. Following media change, the plates were placed back into the incubator until time of transfection. All media replacements involving HEK 293 cells were performed under strict caution as not to dislodge the cells from the cell culture plate surface.

**Table 2.11: Cell line seeding densities for 6-well plates**

<b>Cell Line</b>	<b>Seeding Density (cells per well in a 6 well plate)</b>
<b>HEK 293 Cells</b>	$5 \times 10^5$
<b>Hs 578 T</b>	$3 \times 10^5$
<b>MCF7 Cells</b>	$3 \times 10^5$
<b>MDA MB 231 Cells</b>	$3 \times 10^5$
<b>T47 D Cells</b>	$3 \times 10^5$



Each luciferase reporter gene deletion construct, along with the pRSV $\beta$ gal construct, was transfected into 3 wells, therefore performed in triplicate for every experiment<sup>4</sup>. Each well was transfected with 0.5 $\mu$ g of a specific luciferase reporter gene deletion construct and 0.25 $\mu$ g of the pRSV $\beta$ -gal plasmid using 6 $\mu$ l of plus reagent (Invitrogen) and 6 $\mu$ l of lipofectamine (Invitrogen) according to the manufacturer's protocol. Briefly, sufficient amounts of plus reagent were added to master mixes containing plasmid construct DNA diluted into appropriate amount of cell line-specific SFM, mixed, and incubated at room temperature for 15 minutes. Following this latter incubation, lipofectamine was diluted in ample SFM (800 $\mu$ l per reporter construct included in the master mix) and was added to each master mix. The complexes were then incubated for another 15 minutes at room temperature. Then, additional SFM was added to the master mixes to make up the final volume so that each well received 1ml of the DNA-Plus reagent-lipofectamine complex mixture. One millilitre of these master mixes was then each added to their respective wells and cells were placed into the incubator and left to incubate for 4 hours. Following this 4 hour time period, the DNA-Plus-lipofectamine complexes were aspirated off and replaced by fresh supplemented media corresponding to each respective cell line. Forty-four hours following this time point, cells were lysed and protein was extracted.

---

<sup>4</sup> Depending on the experiment, and along with the luciferase reporter gene constructs and the pRSV $\beta$ -gal expression vector, cells were also transfected with either 0.5 $\mu$ g of either hER $\alpha$  pCDNA3 (for HEK 293 cells) or 0.5 $\mu$ g of pCDNA3 empty vector (for breast cancer cell lines). For further information as to the rationale for the use of these pCDNA3 constructs and as to when exactly these vectors were employed, refer to the results section of this thesis for each experimental data set.

### 2.5.2 MCF7 Cell Transient Transfection

Approximately 18 hours prior to transfection, MCF7 cell lines were seeded at a density of  $3 \times 10^5$  cells/ well in 6-well plates containing 2ml of MCF7 supplemented media per well. Directly preceding transfection, cell media was replaced with 1.5ml of fresh new MCF7 supplemented media and placed back into the incubator until time of transfection. Each luciferase reporter gene was transfected in triplicate with either 1 $\mu$ g or 0.5 $\mu$ g<sup>5</sup> of luciferase reporter gene plasmid DNA, along with 0.25 $\mu$ g of pRSV $\beta$ -gal plasmid DNA<sup>6</sup>. For MCF7 cells, the *TransIT*<sup>®</sup>-LT1 transfection reagent (Mirus; catalogue # MIR 2300) was employed as a transfection reagent and all transfections were performed according to the manufacture's protocol using a ratio of 3 $\mu$ l of reagent per 1 $\mu$ g of plasmid DNA. Briefly, the *TransIT*<sup>®</sup>-LT1 transfection reagent was brought to room temperature and gently mixed prior to use. Master mixes for each construct were prepared by adding ample amounts of luciferase reporter gene deletion construct DNA, *TransIT*<sup>®</sup>-LT1 transfection reagent, and enough MCF7 SFM so that each well obtained 260 $\mu$ l from the respective master mix (MCF7 SFM: DMEM, 1% Na-pyruvate, and 0.001% of 4mg/ml insulin). These components were incubated at room temperature for 18 minutes, and then consecutively 260 $\mu$ l of each respective master mix was added to its respective well. The culture plates were then placed directly in

---

<sup>5</sup> Varying amounts of *mier1* promoter luciferase reporter gene deletion constructs were used depending on the experiment. For example, the characterization of MLP-P1 activity in MCF7 cells employed 1 $\mu$ g of each reporter construct because this was as 1 $\mu$ g of DNA was an efficient amount as stated by the transfection reagent manufacturer (Mirus). Comparison of MCF7 data to other cell lines though required 0.5 $\mu$ g of reporter construct as this was the same amount transfected into the other cell lines.

<sup>6</sup> For experiments whereby MCF7 data was compared to HEK 293 data, 0.5 $\mu$ g of pCDNA3 empty vector was also transfected in along with the luciferase reporter gene deletion constructs and the pRSV $\beta$ -gal expression vector. For further information as to rational behind this inclusion refer to section 3.2.2 and 3.2.3.

the 37°C incubator and left for 48 hours at which time the cells were then lysed and protein was extracted.

## **2.6 Cell Lysis & Protein Extraction**

Forty-eight hours following transfection, cell media was aspirated from each well and the transfected cells were washed with 1ml of 1 x PBS. Following aspiration of the 1 x PBS, protein extraction was performed by adding 400µl of 1 x Cell Culture Lysis Reagent (-prepared by adding 4 volumes of sterile dH<sub>2</sub>O to 1 volume of 5 x Cell Culture Lysis Reagent [Promega; 125mM Tris (pH 7.8 with H<sub>3</sub>PO<sub>4</sub>), 10mM DTT, 50% glycerol, and 5% Triton X-100, stored at -20°C] to each well at room temperature. The plates were then incubated on ice for at least 10 minutes to further enhance cellular lysis. Following this incubation, cells were scraped from each individual well using separate cell lifters (Fisher Scientific) for each respective transient transfection, and lysate was transferred to a chilled 1.7ml eppendorf tube (Corning). Each lysate sample was then vortexed for 10 seconds and then centrifuged for 15 seconds at 12,000g in an Eppendorf Centrifuge model 5415C. Supernatant was transferred to a freshly chilled 1.7ml eppendorf tube, and each sample was stored at -20°C until further use.



## 2.7 Reporter Assays

### 2.7.1 Beta-Galactosidase ( $\beta$ -gal) Assay

Twenty four hours following collection of cell lysate, samples were thawed, vortexed, spun down, and aliquoted into 1.7ml eppendorf tubes in order to proceed with the  $\beta$ -galactosidase ( $\beta$ -gal) assay. This assay was performed in order to analyze the transfection efficiency of each well/sample by measuring the ability of  $\beta$ -gal to convert its substrate, o-nitrophenyl- $\beta$ -D-galactopyranoside, to O-nitrophenol, which proceeds as a calorimetric reaction. The amount of sample aliquoted in preparation for the assay depended on the cell line used and the transfection efficiency of each particular cell line (Refer to Table 2.12). Three separate tubes were additionally prepared as blanks/ negative controls by aliquoting 10 $\mu$ l of 1 x cell culture lysis reagent instead of sample lysate. The assay itself was performed at either room temperature or 37°C, depending on cell line used and the nature of the experiment (Refer to Table 2.12). Two hundred microlitres of  $\beta$ -gal assay buffer [100% Z-buffer (-prepared using protocol from yeast protocol handbook-Clontech: 4g/L of *ortho*-Nitrophenyl- $\beta$ -galactoside (ONPG)- the substrate of  $\beta$ -galactosidase enzyme (Sigma N-1127), and 0.27% of beta-mercaptoethanol ( $\beta$ -M-ETOH) (Bio-Rad)] was added in specific time intervals to each respective lysate sample and then vortexed for 1-2 seconds. The time of when the  $\beta$ -gal buffer was added was recorded and the calorimetric reaction was left to proceed until a visible yellow colour was formed. Once the samples turned yellow, the reaction was stopped by adding 200 $\mu$ l of 1M Tris pH 11.3 stop solution [121.1 g Tris Base, 1 litre dH<sub>2</sub>O, adjusted to pH 11.3 and autoclaved] in the same interval sequence as the lysate and the time of the sample incubation with the  $\beta$ -gal buffer alone was calculated. Two hundred microlitres of

each sample was then transferred to a 96-well plate (Corning) and the absorbance at 415nm was read for each sample using a spectrophotometer. Raw beta gal data was then corrected for and normalized according to the value of the blank readings and time of assay duration by dividing the value of the 415nm sample reading minus the 415nm blank reading by the  $\beta$ -gal incubation time period.

**Table 2.12: Temperatures & sample aliquot amounts for  $\beta$ -Gal assay**

Cell Line	Amount of Lysate Aliquoted ( $\mu$ l)	Temperature of Assay ( $^{\circ}$ C)	
		MLP-P1 Promoter Characterization	MLP-P1 Promoter Comparison between Cell Lines
HEK 293 Cells	5	25	37
MCF7 Cells	10	25	37
MDA MB 231 Cells	10	37	37
T47D Cells	10	37	37
Hs578T Cells	10	37	37



### 2.7.2 Luciferase Assay

Promoter strength, and thus the ability of a promoter to recruit RNA pol II for transcription initiation can be quantified using the luciferase assay system (Promega). This system employs reporter gene vectors harbouring a promoterless gene encoding the firefly luciferase enzyme, of which can be measured upon addition of the luciferase substrate and subsequent generation of light. It was this assay system that was employed throughout the majority of this project. Briefly, cell lysate samples were thawed, vortexed, and quickly spun down in preparation for the luciferase assay the day following protein extraction. Lyophilized luciferase assay substrate (Promega; catalogue # E1501) was then brought to room temperature, for at least a half an hour in an aluminum-covered plastic container, so as to prevent any exposure to light. Once the luciferase assay substrate was equilibrated to room temperature, 10 $\mu$ l of each cell lysate sample was transferred to a luminometer tube (Simport; T405-3) and combined and gently mixed with the substrate for 10 seconds. The amount of light produced following the enzymatic reaction between the luciferase protein and its substrate was read for 10 seconds by a luminometer (Monolight 2010 luminometer; Analytical Luminescence Laboratory) and the relative luciferase units (RLU) values were recorded. RLU values were then normalized to transfection efficiency and the amount of protein in each sample, which was calculated by dividing the RLU values by the ratio of the corrected beta gal values (section 2.7.1) to the Bio-Rad values (section 2.7.3) for each respective sample.

### **2.7.3 Bio-Rad Protein Assay**

Bio-Rad Protein assays were completed in order to determine the amount of protein present in each cell lysate sample to facilitate the normalization of each RLU value to the level of protein of each respective sample. Cell lysate samples were thawed, vortexed, and quickly spun down prior to being aliquoted. Bio-Rad protein assay dye reagent (Bio-Rad) was then equilibrated to room temperature. Following the temperature equilibration of the Bio-Rad protein assay dye reagent, 790 $\mu$ l of sterile dH<sub>2</sub>O was added to 10 $\mu$ l of cell lysate. Two hundred microlitres of Bio-Rad protein assay dye reagent was then added to the diluted cell lysate and vortexed. For a control or blank sample, in order to calibrate the spectrophotometer, 200 $\mu$ l of Bio-Rad dye reagent was added to 790 $\mu$ l of sterile dH<sub>2</sub>O and 10 $\mu$ l of 1 x cell culture lysis buffer. Samples were then incubated for at least 5 minutes at room temperature and then transferred to 1.5ml cuvettes (Fisherbrand). The absorbance of each sample was measured at 595nm following calibration of the spectrophotometer (Pharmacia Biotech Ultrospec 2000).

### **2.8 Immunocytochemistry (ICC) & Transient Transfection**

In order to investigate the nature of the putative NES encoded by exon 3A and its effects on the subcellular localization of MIER1 $\alpha$  in breast cancer cells and in the non-cancerous HEK 293 cell line, immunocytochemistry (ICC) was performed following transient transfection of expression vectors encoding MIER1 $\alpha$  with and without the 23 amino acids encoded by exon 3A. In preparation of ICC, HEK 293, MCF7, T47D, and MDA MB 231 cells were seeded into either poly-L-lysine hydro-



bromide (Sigma) treated 8-well chamber slides (BD BioSciences) for MCF7 cell lines, or 0.1% gelatin (Fisher Scientific) treated 8-well chamber slides (BD BioSciences) for HEK 293 cells and MDA MB 231 cells<sup>7</sup>. All cells lines were seeded into their respective media at the densities outlined in Table 2.13 eighteen hours prior to transfection<sup>8</sup>. Transient transfections with either pCS3+MT empty vector, pCS3+MT*mier1a*, or pCS3+MT *mier1a* exon 3A were performed using *TransIT*<sup>®</sup>-LT1 transfection reagent (for MCF7 & MDA MB 231cells) and *TransIT*<sup>®</sup>-293 transfection reagent (Mirus; catalogue # MIR 2710) for HEK 293 cells. Each procedure followed was exactly the same as described in section 2.5.1 and 2.5.2 respectively, with the exception of the various reagent volumes which were changed to coincide with volumes corresponding to 8-well chamber slides as per manufacturer protocol<sup>9</sup>. Briefly, the cell media of each well of each respective chamber slide was replaced by 200µl of fresh MCF7, MDA MB 231, or HEK 293 supplemented media. Each well was transfected with 0.26µg of expression vector plasmid DNA. *TransIT*<sup>®</sup>-LT1 transfection reagent (*TransIT*<sup>®</sup>-293 transfection reagent for HEK 293 cells) was employed at a 3:1 ratio of reagent to plasmid DNA for MCF7 and HEK 293 cells and a 8:1 ratio for MDA MB 231 cells. Master mixes for each expression vector were prepared so that each chamber slide well received 0.26µg of expression vector DNA,

---

<sup>7</sup> The reason why HEK 293 cells and MDA MB 231 cells were seeded into 0.1% gelatin treated chamber slides instead of the poly-l-lysine hydro-bromide treated slides used for the MCF7 cells was because the 0.1% gelatin treated chamber slides more efficiently kept the cells of these respective cell lines remaining on the chamber slide surface. Unlike MCF7 cells, which are readily adherent to the chamber slide surface following poly-l-lysine hydro-bromide treatment, the HEK 293 and MDA MB 231 cells effortlessly dislodge from the surface when poly-l-lysine hydro-bromide is applied as a cell-adherent agent.

<sup>8</sup> Table 2.15 illustrates the optimal seeding densities for 8-well chamber slides of each cell line as dictated by the distinct growth patterns and cell structure of each respective cell line.

<sup>9</sup> The protocol employed for the *TransIT*<sup>®</sup>-293 transfection reagent was the same as that which was used for the *TransIT*<sup>®</sup>-LT1 transfection reagent.



0.8µl of *TransIT*<sup>®</sup> transfection reagent, and 24.68 µl of cell line-specific SFM. The master mixes were incubated at room temperature for 18 minutes, and then consecutively, 26µl of each respective master mix was added to the coinciding wells. The culture plates were then placed directly in the 37°C incubator.

Forty-eight hours following transient transfection, the chamber slides were removed from the incubator and placed in a 250ml beaker of 1 x PBS (diluted from 10 x PBS: 80g NaCl, 2g KCl (MP Biomedicals, LLC), 2.4g KH<sub>2</sub>PO<sub>4</sub> (Fisher Scientific), 21.6g Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O (Fisher Scientific), adjusted to pH 7.4, final volume to 1 litre with dH<sub>2</sub>O, autoclaved) in order to wash and rinse the supplemented media from the cells. Cells were then fixed to the chamber slides by first removing the chamber slides from the 250ml beakers containing 1 x PBS so that each well was full with 1 x PBS. Half of the 1 x PBS was then aspirated and then 200µl of 4% paraformaldehyde solution [2g paraformaldehyde (Fisher Scientific), 5ml 10 x PBS, 40mL dH<sub>2</sub>O warmed to 60°C in order to dissolve paraformaldehyde, adjusted to pH 7] to each chamber slide well and incubated for 5 minutes. Chamber slides were then again placed into 250ml beakers of 1 x PBS for rinsing and then removed so that 1 x PBS solution remained in each well. Half of the 1 x PBS was then again aspirated and 200µl of 4% paraformaldehyde was added to each well and incubated for 10 minutes in a moist incubation or humidification chamber in order to prevent the cells from drying during incubation thereby allowing a total paraformaldehyde-incubation time of 15 minutes. The humidification chamber employed for these steps and future steps of the ICC protocol was prepared by inverting a small plastic gel tray and placing it in a larger sterile plastic container lined with paper towel napkins soaked in 1 x PBS.

Chamber slides were then washed twice in fresh 1xPBS by submersing them in two fresh 250ml 1 x PBS filled beakers. The chamber slides were then placed in 250ml beakers containing 0.1% TritonX-100/ 1 x PBS [100% TritonX-100 (Sigma) diluted in 1 x PBS] for 5 minutes in order to disrupt and perforate the cellular membrane. This step strengthens the access and subsequent recognition and staining of the antibody-specific antigens. The 0.1% TritonX-100/ 1 x PBS was then aspirated off and 200µl of 5% blocking buffer was then added to each well and the chamber slides were incubated for 1 hour in a humidification chamber. This step is essential in the ICC procedure as the blocking buffer binds to the non-specific sites inside the cell that could be recognized by the secondary antibody and therefore prevents future non-specific staining from occurring. The blocking buffer is able to function in such a manner as it is made up of serum from the same animal source to which the secondary antibody was prepared. In the case of the experiments performed in this study, the secondary antibody, sheep-anti mouse horseradish peroxidase (SAM-HRP) (Amersham Bioscience), was engineered from immunization of a sheep against the mouse heavy chain antigen incorporated in the primary antibody that is of murine origin. Therefore, the 5% blocking buffer constituted 5% sheep serum in 1 x PBS. Following the 1 hour blocking buffer incubation, chamber slides were submersed in a 250ml beaker containing 1 x PBS and then, following aspiration of the 1 x PBS, 200µl of 0.6% H<sub>2</sub>O<sub>2</sub> (Fisher Scientific) was added to each well and incubated for 30 minutes in a humidification chamber. Hydrogen peroxide treatment of the cells is essential as it degrades any endogenous peroxidases that would otherwise interact with the later employed secondary antibody reaction with SIGMA FAST™ 3, 3'- Diaminobenzidine (DAB) tablets (Sigma) and produce non-specific staining.



Following the latter incubatory period, chamber slides were washed in 1L beakers containing 1 x PBS. Two hundred microlitres of primary antibody diluted at the optimal dilution determined for ICC of 1: 200 in 3% Bovine Serum Albumin (BSA) RIA grade (Sigma)/ 1 x PBS was added to each respective chamber well following the aspiration of the 1 x PBS. The primary antibody used in this study is a monoclonal antibody that was derived from murine hybridoma cells produced from the fusion of murine myeloma cells with the murine spleen cells from a mouse that had been immunized with the MYC tag epitope antigen, which is incorporated in the proteins expressed from the expression vectors used in this study: pCS3+MT empty vector, pCS3+MT*mier1a*, and pCS3+MT *mier1a* Exon 3A. This antibody will thus recognize and bind the specific proteins encoded by the transfected expression vectors in order to determine the subcellular localization of these specific *mier1* protein isoforms. The chamber slides were then left to incubate over night in a humidification chamber wrapped in paraflim at 4°C. At least one well for each chamber slide was used as a control containing no primary antibody, whereby 200µl of only 3% BSA/ 1 x PBS was added. The non-primary antibody containing well is pivotal for performing ICC as it allows the investigator to ensure that the resulting final staining is due to the secondary antibody binding to the primary antibody and not due to non-specific binding of the secondary antibody.

The next day, the primary antibody was then aspirated off and each chamber slide was dipped into a 1L beaker containing 0.1% TritonX-100/ 1 x PBS and were then removed and incubated in 250ml beaker containing 1 x PBS for 5 minutes. Following aspiration of the 1 x PBS from each chamber slide well, the secondary antibody (SAM-HRP), specific for the primary antibody used, was diluted 1:200 in



3% BSA / 1 x PBS and was added and incubated for 1 hour at room temperature in a humidification chamber. After the 1 hour time point, chamber slides were then again dipped into the 1L beaker containing 0.1% TritonX-100/ 1 x PBS and subsequently incubated in 250ml beaker containing 1 x PBS for 5 minutes. 1 x PBS was then aspirated and 200 $\mu$ l DAB (Sigma) solution, which was prepared as per manufacturer's protocol, was added to each well. The DAB reacts with the horseradish peroxidase that is conjugated to the secondary antibody thereby producing a reddish- brown precipitate exactly where the secondary antibody binds to the primary antibody recognizing the protein antigen of interest. This reaction was left to proceed till staining was visible (approximately 15-30 minutes). DAB solution was then aspirated off each well and the chamber slides were submersed in a 1L beaker containing fresh 1 x PBS and incubated for 5 minutes. The gasket was then removed following aspiration of the 1 x PBS and 2 drops of 10% glycerol were gently added to each well. Slides were then mounted, a cover slip was added, and each slide was sealed with clear nail polish. Slides were then viewed and analyzed with a compound light microscope (Olympus) under the 10X, and 20X objective using both bright field and phase contrast microscopy. Stained cells were counted, and the percentage of nuclear, cytoplasmic, and whole cell staining for each experimental condition and chamber slide well was recorded. Specifically, staining was considered nuclear if only the nucleus was exclusively stained. Consequently, the cytoplasmic category consisted of cells that exclusively displayed cytoplasmic staining, where as the whole cell staining category constituted cells that contained both nuclear and cytoplasmic staining at approximately equal levels. Pictures of representative field of views were taken using

the Cool Snap system version 1.1 (Roper Scientific). Slides were stored at 4°C in the dark in order to avoid bleaching and degradation of the staining.

**Table 2.13: Cell line seeding densities for 8 well-chamber slides**

<b>Cell Line</b>	<b>Seeding Density (cells per well in an 8 well chamber slide)</b>
<b>HEK 293 Cells</b>	12, 000 cells/ well
<b>MDA MB 231 Cells</b>	15, 000 cells/ well
<b>MCF7 Cells</b>	15, 000 cells/ well



### 2.8.1 Leptomycin B Treatment

Leptomycin B (LMB) (Alexis Biochemicals) was added to transfected MCF7 cells at a concentration of 10ng/ml in order to inhibit nuclear export. Non LMB treated transfected MCF7 cells were also used as negative controls whereby the LMB vehicle, 0.1% ethanol, was added in equal volume. LMB treatment was employed 40 hours post transfection for 8 hours.

### 2.9 Statistical & *in silico* Analyses

All statistical analyses were performed using the INSTAT version 3.0 software program statistical program (GraphPad Software, San Diego, CA, USA) using the Student's t-test,  $p < 0.01$ . All data values analyzed were assumed to be sampled from Gaussian distributions using standard parametric methods.

As briefly mentioned in section 2.4.1, all *in silico* analyses investigating putative TFBS present at specific regions of the MLP-P1 promoter were performed using algorithm tools on four separate websites (refer to section 2.4.1). In the case of the Transcription Factor Binding Site search system, only TFBSs with a score of at least 80% were finally considered as a putative TFBS. The algorithm system 'consite' was employed using a "bit" parameter of 10 for its search. The 'bit' parameter measures specificity, whereby the higher the number, the more stringent the results. Mathematically, two bits refer to a precisely specified individual base. Therefore, a 10 bit threshold would find an exact match to a 5 base site (Sandelin *et al.*, 2004).

Furthermore, the percentage of homology of between the human MLP-P1 promoter nucleotide sequences involved in minimal and maximal promoter activity as well as the nucleotide sequences upstream of and encompassed with exon 3A were compared to six different species [rhesus (monkey), dog, horse, mouse, rat, and chicken] using the VISTA Browser from the VISTA Comparative genomics website (<http://genome.lbl.gov/vista/index.shtml>). This specific program measures conserved regions of DNA from various pre-computed genome alignments of different species using multiple alignment, visualization, and statistical analysis tools (Brudno *et al.*, 2007; Frazer *et al.*, 2004)

## Chapter 3- Results

### 3.1 Characterization of the MLP-P1 Promoter Region of *mier1* in Breast Cancer Cells

Recent studies have implicated a role for MIER1 $\alpha$  in breast cancer, not only because this protein is a corepressor of ER $\alpha$ , but also because its overexpression inhibited estrogen-induced anchorage independent growth of an ER+ breast cancer cell line (Mercer *et al.*, 2008; McCarthy *et al.*, 2008). Moreover, MIER1 $\alpha$  may be a candidate contributing factor in invasive breast cancer as this protein is increasingly abnormally localized to the cytoplasm when tumours progress to a more invasive state. One hypothesis is that this change in subcellular localization may possibly enhance the progression of invasive breast cancer by interfering with or abrogating MIER1 $\alpha$ 's potent transcriptional regulatory role in the cell, with the suppression of ER $\alpha$  being one of its more influential functions with regards to breast tumourigenesis (refer to section 1.5.5 and 1.5.3.2.3).

As briefly alluded to in section 1.5.6, a possible mechanism for this documented change in subcellular localization could be preferential or increased transcription initiation at the MLP-P1 promoter, which potentially could result in proteins containing a putative NES and/ or transmembrane domain encoded by the MLP-P1 specific exon 3A. Therefore, in order to investigate whether abnormal transcriptional regulation of *mier1*'s MLP-P1 promoter is apparent in breast cancer, the promoter activity stemming from both the proximal promoter regions, as well as further upstream sequences of the MLP-P1 promoter region, were characterized in breast cancer cells.



The proximal promoter of a gene, which typically lies within 250 bp upstream of the TSS, is the area of the promoter that contains elements critical to transcriptional regulation (Heintzman & Ren, 2007). Preliminary data investigating the MLP-P1 promoter region from our lab demonstrated that a stretch of nucleic acids residing between 185bp upstream and 37bp downstream of the putative MLP-P1 TSS (1463bp and 1241bp upstream of the MLP-P1 translation start site) was involved in eliciting activation at this promoter (refer to Appendix 4)<sup>10</sup>. In the case of this study, putative MLP-P1 promoter regions of interest were inserted into the luciferase reporter gene plasmid upstream of the luciferase gene so as to act as the luciferase promoter. These luciferase reporter gene constructs essentially divided the aforementioned stretch of MLP-P1 putative proximal promoter DNA sequence between -185bp and +37 bp into three distinct parts, with MLP-P1 (-185) pGL3 containing the entire sequence downstream of -185 to seven nucleotides preceding the ATG translation start site of MLP-P1 and MLP-P1 (+37) pGL3 only containing the entire sequence downstream of +37 up to seven nucleotides preceding the MLP-P1 ATG translation start site (Fig 2.1 & 2.2). In order to divide the sequence between -185 and +37 into further sections for a more thorough analysis, two further deletion constructs were engineered as described in section 2.4<sup>11</sup>:

---

<sup>10</sup> As the actual transcription start site of MLP-P1 has not yet been officially characterized, all nucleotide counts are indicated with reference to the MLP-P1 ATG translation start site. The first exon (exon 1A) encoded by MLP-P1 lies 1279bp upstream of the MLP-P1 ATG start site, and therefore the putative TSS of this promoter is hypothesized to lie at this position. Therefore with reference to the putative TSS of MLP-P1, this particular stretch of amino acids that was primarily analyzed extends from 185bp upstream to 37bp downstream of the MLP-P1 putative TSS.

<sup>11</sup> Refer to section 2.4 for further details on cloning, construction, and the exact sequences inserted into each respective luciferase reporter gene deletion construct.

1. **MLP-P1(-91)**; the sequence of which was cloned into the pGL3 luciferase plasmid vector starts 94bp downstream of the MLP-P1 (-185) pGL3 inserted sequence and 125bp upstream of the MLP-P1 (+37) pGL3 inserted sequence.
2. **MLP-P1 (-44)**; the sequence of which was cloned into the pGL3 luciferase plasmid vector starts 141bp downstream of the MLP-P1 (-185) pGL3 inserted sequence and 81bp upstream of the MLP-P1 (+37) pGL3 inserted sequence.

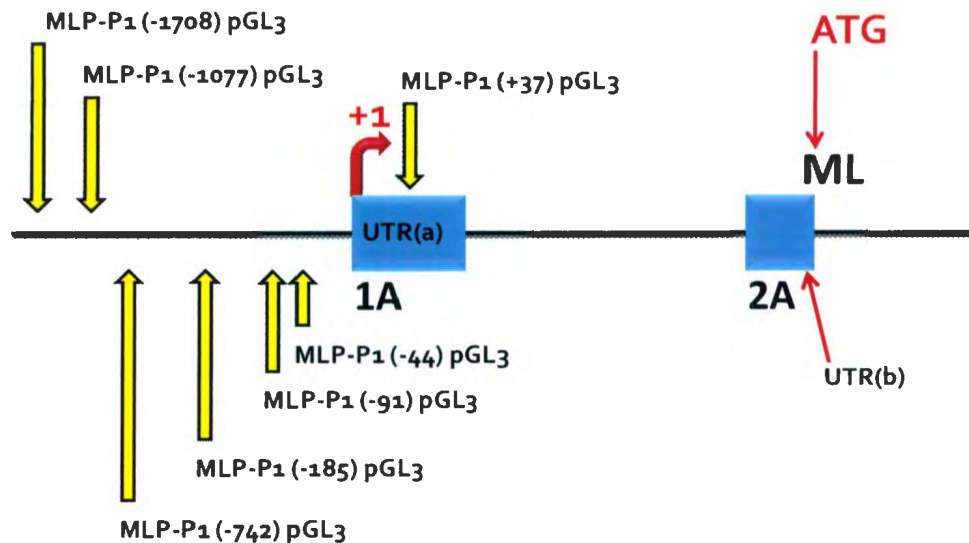
Furthermore, in addition to these four aforementioned MLP-P1 luciferase reporter gene deletion constructs, the following three constructs were used: MLP-P1 (-1708) pGL3, MLP-P1 (-1077) pGL3, and MLP-P1 (-742) pGL3<sup>12</sup> (Fig 3.1). The reason for this additional incorporation was to include sequences further upstream of the MLP-P1 putative proximal promoter region in order to investigate whether *mier1* underwent regulation by distal regulatory elements, such as enhancers or silencers, which have been known to exert their regulatory activity considerable distances away from proximal promoter regions (Maston *et al.*, 2006). Acquisition of information with regards to such long range *cis* transcriptional regulatory elements would supply additional pertinent information with regards to the nature of MLP-P1 regulation in breast cancer.

One microgram of each of these MLP-P1 luciferase reporter gene deletion constructs (depicted in Figure 3.1 and furthermore described in section 2.2.3.1), 1 µg of the pGL3 empty vector negative control, and 0.25µg of the pRSVβ-gal expression vector were each transiently transfected into MCF7 cells as described in section 2.5.2.

<sup>12</sup> The bracketed numbers denoting these luciferase reporter gene deletion constructs indicate how many base pairs upstream or downstream the 5' ends of the insertion sequences inserted into each respective construct lies with respect to the MLP-P1 putative TSS. For further information about the location of these insertion sequences with relation to the MLP-P1 promoter region refer to section 2.2.3.1 & appendix 1.

MCF7 cells express ER $\alpha$  and have long been used as a standard model for breast cancer (Kern *et al.*, 2004; Levenson & Jordan, 1997). For our purposes, they served as a suitable model to begin the MLP-P1 promoter activity analysis. Forty-eight hours following transfection, cells were harvested and cell lysates were collected. Relative luciferase units (RLUs) were then measured using the luciferase assay protocol as described in section 2.7.2. RLUs of each sample were normalized to  $\beta$ -gal assays values, which measured transfection efficiency, as well as to the protein level readings obtained following experiment-specific bio-rad. Bio-rad assays took into consideration the amount of protein in each sample which reflects the number of cells collected.





**Figure 3.1: Schematic of luciferase reporter gene deletion constructs used for the characterization of the minimum and maximum promoter activity regions of MLP-P1 in MCF7 breast cancer cells**

This schematic depicts the MLP-P1 promoter region and the 5' starting position of the MLP-P1 putative promoter region inserts cloned into each individual MLP-P1 luciferase reporter gene deletion construct that was used in section 3.1. Additional notation depicted in this schematic describes the MLP-P1 proximal promoter region which is further described in section 2.2.3.1.

### **3.1.1 The Minimal Promoter of MLP-P1 in MCF7 Breast Cancer Cells**

Figure 3.2 illustrates results from five independent experiments in MCF7 cells, which recapitulated preliminary findings that the region between -185 and +37 does encompass an active promoter region driving transcription at MLP-P1, as MLP-P1 (-185) pGL3 and MLP-P1 (-91) pGL3 generated a significant increase in RLU production over the pGL3 empty vector ( $p < 0.001$  and  $p < 0.05$  respectively). MLP-P1 (+37) pGL3 though, as well as the construct harbouring the second smallest insert, MLP-P1 (-44) pGL3, did not produce a significant increase above the RLU level of the pGL3 empty vector negative control ( $p > 0.05$ ) (Fig 3.2). These results denote MLP-P1(-91) pGL3 as the MLP-P1 luciferase reporter gene deletion construct containing the smallest sequence able to produce a significantly higher level of RLU over that of the pGL3 empty vector in MCF7 cells (Fig 3.2).

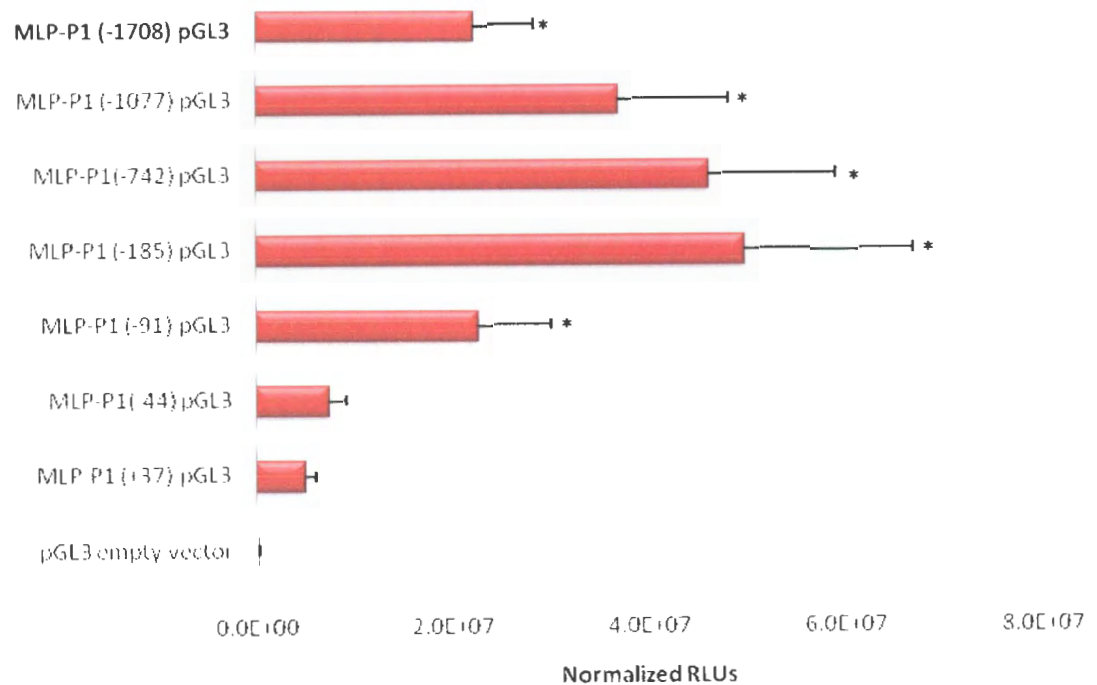
### **3.1.2 Maximal Promoter Activity at MLP-P1 in MCF7 Breast Cancer Cells**

As clearly illustrated in Figure 3.2, the MLP-P1 (-185) pGL3 construct generated the highest level of activity above the levels of the pGL3 empty vector negative control ( $p < 0.001$ ). Even the larger MLP-P1 luciferase reporter gene deletion constructs, which harboured further distal upstream MLP-P1 sequences, as well as the same nucleic residues inserted in MLP-P1 (-185) pGL3, did not produce significantly higher luciferase activity levels than the MLP-P1 (-185) pGL3 construct ( $p > 0.05$ ). This marks MLP-P1 (-185) pGL3 as the construct with the smallest insert resulting in the highest promoter activity. The rest of the constructs in the deletion series that

harbour smaller MLP-P1 sequence inserts revealed significantly lower luciferase activity than MLP-P1 (-185) pGL3, which is succinctly exemplified following analysis of the MLP-P1 (-91) pGL3 minimal promoter construct. This construct, which interestingly only contains 94 less nucleotide base pairs than MLP-P1 (-185) pGL3 significantly decreases in luciferase activity production by 2 fold ( $p < 0.01$ ) when compared to MLP-P1 (-185) pGL3 (Fig 3.2).

The stretches of *mier1* MLP-P1 sequence upstream of -185 did not produce higher levels of promoter activity than MLP-P1 (-185) pGL3. Figure 3.2 demonstrates that the promoter activity stemming from the construct harbouring the longest MLP-P1 putative promoter insert, MLP-P1 (-1708) pGL3, generated levels of luciferase activity that were statistically lower than that of MLP-P1 (-185) pGL3 (Fig 3.2). The p-value upon comparison of these two latter constructs was smaller than 0.001. Furthermore, MLP-P1 (-1708) pGL3 displayed approximately the same promoter activity level as the minimal promoter construct MLP-P1 (-91) pGL3 ( $p > 0.05$ ). MLP-P1 (-742) pGL3 and MLP-P1 (-1077) pGL3 also did not produce higher activity levels than MLP-P1 (-185) pGL3, but rather exhibited statistically comparable reporter activity levels to this latter construct ( $p > 0.05$ ) (Fig 3.2).





**Figure 3.2: Characterization of the MLP-P1 promoter region of *mier1* in MCF7 cells**

MCF7 cells were seeded at a density of  $3 \times 10^5$  cells/well of a 6-well plate and grown in DMEM for approximately 18 hours. Cells were then transfected with 1 $\mu$ g of MLP-P1 luciferase reporter gene deletion construct and 0.25 $\mu$ g of pRSV $\beta$ -gal. Cell lysates were collected 48 hours following transfection and relative luciferase units were measured. BioRad and  $\beta$ -gal assays were performed so as to collect necessary data in order to normalize the RLUs to the transfection efficiency as the amount of protein per sample as described in sections 2.7. Each luciferase reporter gene construct was transfected into triplicate wells and the above data constitute a n=5. Statistically significant differences between specific constructs compared to the pGL3 empty vector are denoted by an asterisk whereby  $p < 0.001$  for the MLP-P1(-1077) pGL3, MLP-P1(-742) pGL3, and MLP-P1(-185) pGL3 luciferase reporter gene deletion constructs,  $p < 0.05$  for the MLP-P1 (-1708) pGL3 and MLP-P1(-91) pGL3 reporter constructs.

### **3.2 Comparison of MLP-P1 Promoter Activity Patterns Across Multiple Breast Cancer Cell Lines Varying in ER Status and the Non-Cancerous Cell Line HEK 293**

#### **3.2.1 MLP-P1 Promoter Activity Patterns in HEK 293 Cells**

To demonstrate MLP-P1 promoter activity from the proximal promoter region of MLP-P1 in a non-neoplastic environment, HEK 293 cells were transfected with the pGL3 empty vector negative control and the following MLP-P1 luciferase reporter gene deletion constructs: MLP-P1 (-185) pGL3, MLP-P1 (-91) pGL3, MLP-P1 (-44) pGL3, and MLP-P1 (+37) pGL3<sup>13</sup>. This human embryonic kidney cell line transfects very efficiently and, even though was primarily established by transfection with adenovirus (Graham & Smiley, 1977), ideally served for our purposes as a non-cancerous control for comparison to breast cancer cells.

It is also important to note that one of the key differences between the HEK 293 cell line and the MCF7 cell line, apart from the neoplastic nature of MCF7 cells, is estrogen signalling. MCF7 cells are a well characterized ER+ cell line, and therefore consistently express the ER $\alpha$ . HEK 293 cells, on the other hand, do not express endogenous ER $\alpha$ . Therefore, in order to verify that any plausible differential in promoter activity between the two cell lines was not affected by a lack of ER $\alpha$  expression, HEK 293 cells were also always transfected with 0.5 $\mu$ g of the ER $\alpha$ -expressing vector pCDNA3<sup>14</sup>, along with 0.5 $\mu$ g of MLP-P1 reporter gene deletion

---

<sup>13</sup> The location of the MLP-P1 putative promoter insert that was cloned into each of these specific luciferase reporter gene deletion constructs with respect to the MLP-P1 proximal promoter region is depicted in figures 2.1, 2.2, and 3.1.

<sup>14</sup> As briefly mentioned in the footnote of section 3.1.1, an additional reason why hER $\alpha$  pCDNA3 was used in the HEK 293 experiments was because the data obtained represent results extracted from the negative control groups

construct and 0.25 $\mu$ g pRSV $\beta$ -gal. Consequently, all luciferase experimental data obtained using HEK 293 cells reflects not only transient co-transfection with MLP-P1 reporter gene deletion constructs and pRSV $\beta$ -gal, but also reflects co-transfection with a vector expressing ER $\alpha$  as well. Also, it is important to note that for the purposes of comparing data between HEK 293 cells and the breast cancer cell lines (such as: MCF7, T47D, and Hs578Ts), all the breast cancer cell samples that were used in each respective comparison were also co-transfected with 0.5 $\mu$ g of the empty vector pCDNA3 in addition to the other constructs. By transfecting the pCDNA3 empty vector into the rest of the examined cell lines, all cell line data could be compared equally as this step alleviated the variability that would have been created had only one cell line been transfected with the pCDNA3 vector.

Fifty-two hours <sup>15</sup> following transfection, cells were harvested and cell lysates were collected as described in section 2.6. Relative luciferase units (RLUs) were then measured and RLUs were normalized to transfection efficiency and protein levels. Figure 3.3 depicts results of seven independent experiments using HEK 293 cells and recapitulated the previously established MLP-P1 promoter activity pattern observed in MCF7 breast cancer cells whereby MLP-P1 (-185) pGL3 produces a statistically higher luciferase activity level than all other constructs (Fig 3.3). Furthermore, the MLP-P1 (-91) pGL3 reporter activity levels were significantly lower than that of MLP-P1 (-185) pGL3 ( $p < 0.001$ ), consistent with previous observations in which the

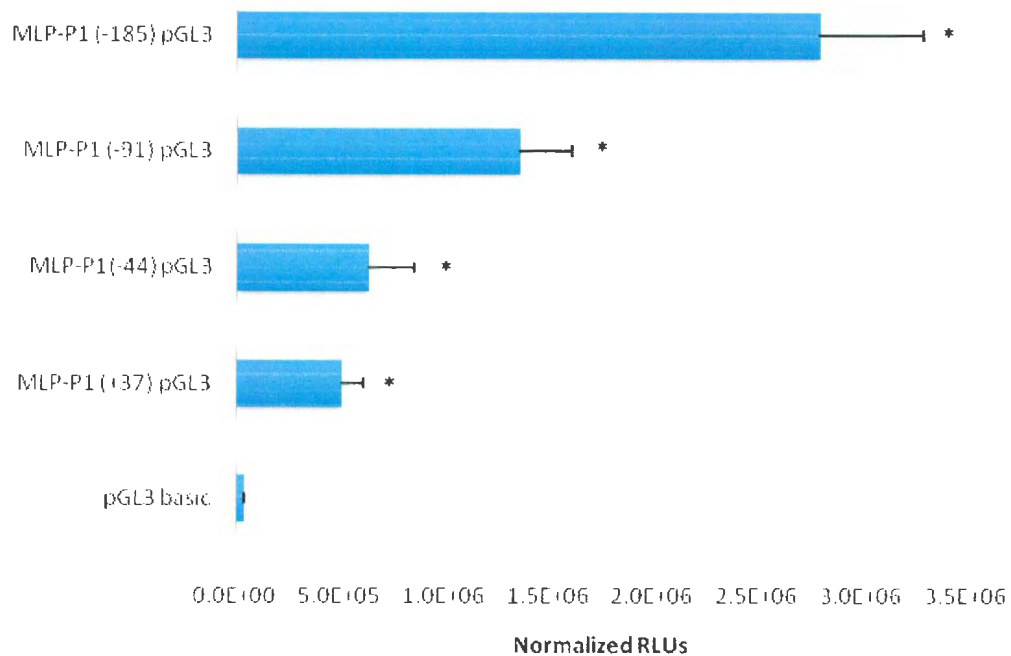
---

of experiments initially investigating the effect of estrogen on MLP-P1 promoter activity (Refer to sections 3.2.2 and 3.3.1.1).

<sup>15</sup> Cells were left to grow for 52 hours because originally these experiments were looking at estrogen treatment effects on MLP-P1 promoter activity. Estrogen treatment was performed 4 hours following transfection and both estrogen-treated and non-treated cells had to be left for 48 hours in the presence of estrogen or treatment vehicle. The results presented in this section of the thesis are data from the non-estrogen treated cells.



94bp MLP-P1 promoter region between -185 and -91 harboured sequences which caused maximal reporter activation. Moreover, RLU values produced by the other MLP-P1 reporter gene deletion constructs sequentially decreased in luciferase activity levels from construct to construct coinciding with the length of the promoter region sequence insert of each MLP-P1 luciferase reporter. For example, the smallest insert, found in the MLP-P1 (+37) pGL3 construct, resulted in the lowest RLU values apart from the pGL3 empty vector (Fig 3.3). This pattern was also consistent with the MLP-P1 promoter characterization in MCF7 cells (Fig 3.2).



**Figure 3.3: Characterization of the MLP-P1 proximal promoter region of *mier1* in HEK 293 cells**

HEK 293 cells were seeded at a density of  $5 \times 10^5$  cells/well of a 6-well plate and grown in DMEM for approximately 18 hours. Cells were then transfected with 0.5  $\mu$ g of MLP-P1 promoter sequence luciferase reporter gene deletion construct, 0.25  $\mu$ g of pRSV $\beta$ -gal, and 0.5  $\mu$ g ER $\alpha$  pCDNA3. Cell lysates were collected 52 hours following transfection and relative light units were measured. BioRad and  $\beta$ -gal assays were performed so as to collect necessary data in order to normalize the RLUs to sample specific transfection efficiency and protein level as described in section 2.7. Each luciferase reporter gene construct was performed in triplicate and  $n=7$ . Statistically significant differences between all constructs compared to the pGL3 empty vector are denoted by an asterisk whereby  $p<0.001$  for MLP-P1(-185) pGL3 and MLP-P1 (-91) pGL3 and  $p<0.05$  for MLP-P1 (-44) pGL3 and MLP-P1(+37) pGL3. Furthermore, all differences in luciferase activity between each respective construct were statistically significant with the exception of MLP-P1(-44) pGL3 vs. MLP-P1 (+37) pGL3. P values for these differences are as follows: a) MLP-P1 (-185) pGL3 vs. all other constructs resulted in  $p<0.001$  b) MLP-P1(-91) pGL3 vs MLP-P1 (-44) pGL3 and MLP-P1(+37) pGL3 resulted in  $p<0.01$ .

### 3.2.2 MLP-P1 Promoter Activity Patterns in Multiple Breast Cancer Cell Lines Varying in ER Status

The estrogen receptor (ER) is presently the most powerful predictive marker in breast cancer management, and its pivotal role in breast tumourigenesis has been, and currently is, vigorously researched (Chen *et al.*, 2008; Payne *et al.*, 2008). MIER1 $\alpha$  has been implicated in ER signalling as a corepressor of ER $\alpha$ , and as an inhibitor of estrogen stimulated anchorage independent growth in a breast cancer cell line following its overexpression (McCarthy *et al.*, 2008). This study further investigated how the MLP-P1 promoter is regulated in breast cancer by examining whether differing ER statuses, and consequently the neoplastic environment developed under the influence of such ER statuses, affect the transcription promoting ability of this promoter and its activity patterns in breast cancer cells.

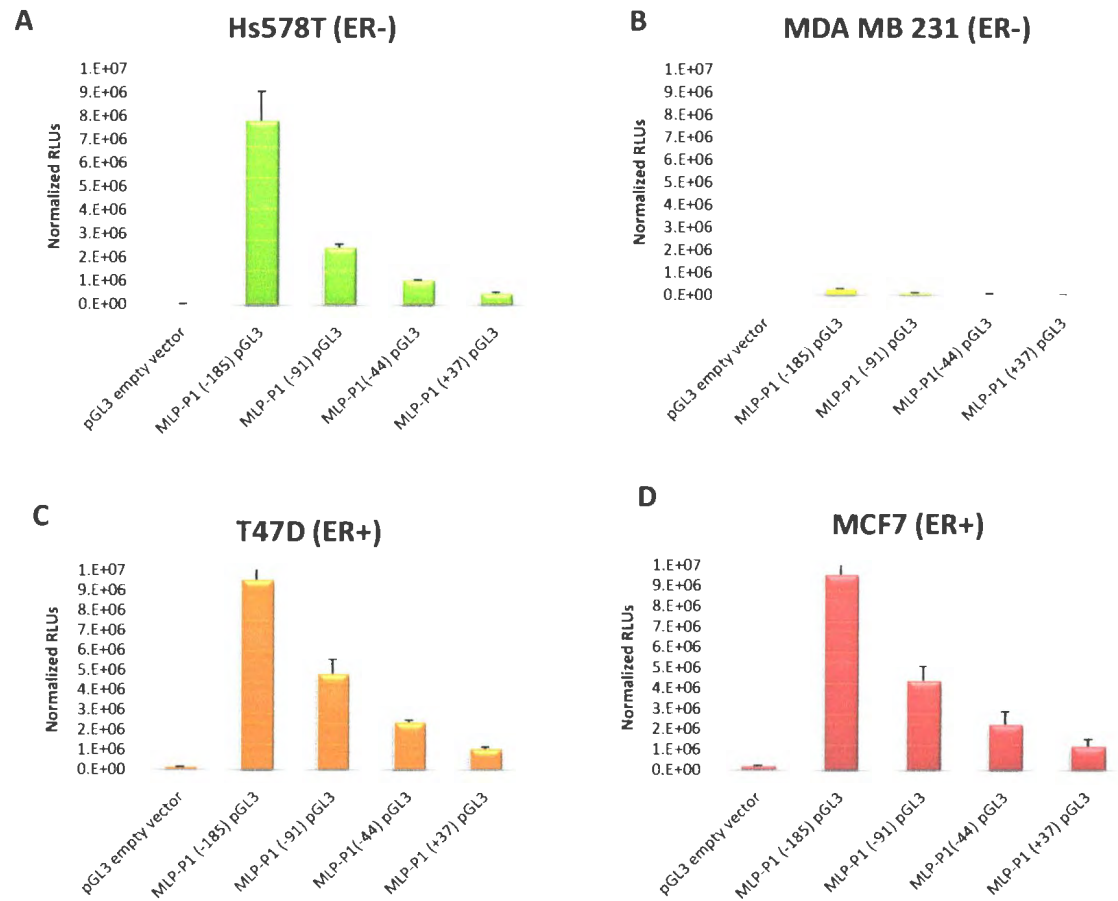
Three additional breast cancer cell lines varying in ER status [T47D (ER+), MDA MB 231 (ER-), and Hs578T (ER-)] along with the MCF7 ER+ breast cancer cell line were transiently transfected with 0.5 $\mu$ g of MLP-P1 luciferase reporter gene deletion constructs and 0.25 $\mu$ g of pRSV $\beta$ -gal expression vector. The MLP-P1 luciferase reporter gene deletion constructs used in this section of this study were the specific constructs that harboured MLP-P1 proximal promoter sequences: MLP-P1 (-185) pGL3, MLP-P1 (-91) pGL3, MLP-P1 (-44) pGL3, and MLP-P1 (+37) pGL3<sup>16</sup>. Furthermore, in order to later compare this data with the previous HEK 293 data, these cell lines were also transfected with the empty vector pCDNA3 (refer to section 3.2.1 for further explanation). All transfections, cell lysis procedures, and reporter assay

<sup>16</sup> The location of the MLP-P1 putative promoter insert that was cloned into each of these specific luciferase reporter gene deletion constructs with respect to the MLP-P1 proximal promoter region is depicted in figures 2.1, 2.2, and 3.1



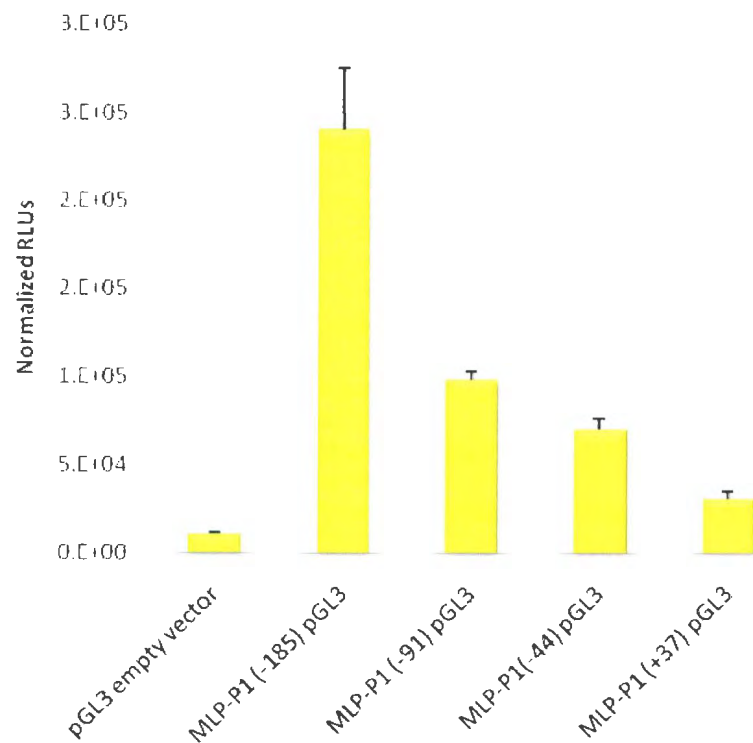
protocols were performed as previously described, and all luciferase values were normalized both to transfection efficiency and protein levels.

Figures 3.4 A, B, C, & D individually depict results stemming from each of the four breast cancer cell lines examined. These data demonstrate that the promoter activity consecutively decreased similarly in each cell line from the maximum activity-producing construct as the MLP-P1 proximal promoter region inserts decrease in length from construct to construct (Fig 3.4 A -E). Furthermore, MLP-P1 (-185) pGL3 persistently produced the highest reporter activity level above the pGL3 empty vector negative control ( $p < 0.001$  for each cell line). Moreover, the luciferase activity generated by the second largest deletion construct, MLP-P1 (-91) pGL3, significantly decreased RLU levels when compared to MLP-P1 (-185) pGL3 in each breast cancer cell line ( $p < 0.001$ ).



**Figure 3.4: MLP-P1 promoter activity across multiple breast cancer cell lines varying in ER status**

MDA MB 231 cells, Hs578T cells, T47D cells and MCF7 cells were seeded at a density of  $3 \times 10^5$  cells/well and in a 6-well plate and grown in cell type specific supplemented DMEM for approximately 18 h. Respective wells were then transfected with 0.5 $\mu$ g of MLP-P1 luciferase reporter gene deletion construct, 0.5 $\mu$ g pCDNA3 empty vector, and 0.25 $\mu$ g of pRSV $\beta$ -gal. Cell lysates were collected 48 hours following transfection, and relative luciferase units were measured. BioRad and  $\beta$ -gal assays were performed so as to collect necessary data in order to normalize the RLUs to transfection efficiency as well as protein levels as described in section 2.7. Each luciferase reporter gene construct was transfected into triplicate wells and the above data constitute: n= 4 (MDA MB 231 cells), n=3 (T47D cells), n=3 (Hs578T cells), and n=5 (MCF7 cells). Figure 3.4A represents relative luciferase units plotted with respect to the MLP-P1 luciferase reporter gene deletion construct data on the X axis for Hs578T cells whereas Figure 3.4B depicts such results for the MDA MB 231 cells (please refer to Figure 3.5 which contains a different y axis scale in order to further distinguish actual RLU values for this particular cell line), Figure 3.4C represents the T47D cell line, and Figure 3.4D represents data from the MCF7 cell experiments.



**Figure 3.5: MLP-P1 promoter activity in MDA MB 231 cells**

MDA MB 231 cells were seeded at a density of  $3 \times 10^5$  cells/well and in a 6-well plate and grown in cell type specific supplemented DMEM for approximately 18 hours. Respective wells were then transfected with 0.5 $\mu$ g of MLP-P1 luciferase reporter gene deletion construct, 0.5 $\mu$ g pCDNA3 empty vector, and 0.25 $\mu$ g of pRSV $\beta$ -gal. Cell lysates were collected 48 hours following transfection, and relative luciferase units were measured. BioRad and  $\beta$ -gal assays were performed so as to collect necessary data in order to normalize the RLUs to transfection efficiency as well as protein levels as described in section 2.7. Each luciferase reporter gene construct was transfected into triplicate wells and the above data constitute: n= 4 (MDA MB 231 cells). This figure represents the exact same data illustrated in Figure 3.4 B, however the values on Y axis are lower so that the MDA MB 231 relative luciferase units are more easily observable. Due to the low RLU values obtained in this specific cell line, when compared to scale with the other breast cancer cell lines, it is difficult to see the MLP-P1 promoter activity pattern. As in Figure 3.4, the RLUs are plotted with respect to the MLP-P1 luciferase reporter gene deletion construct data on the X axis.



### 3.2.3 MLP-P1 Promoter Activity Comparison in HEK 293 Cells and Multiple Breast Cancer Cell Lines Varying in ER Status

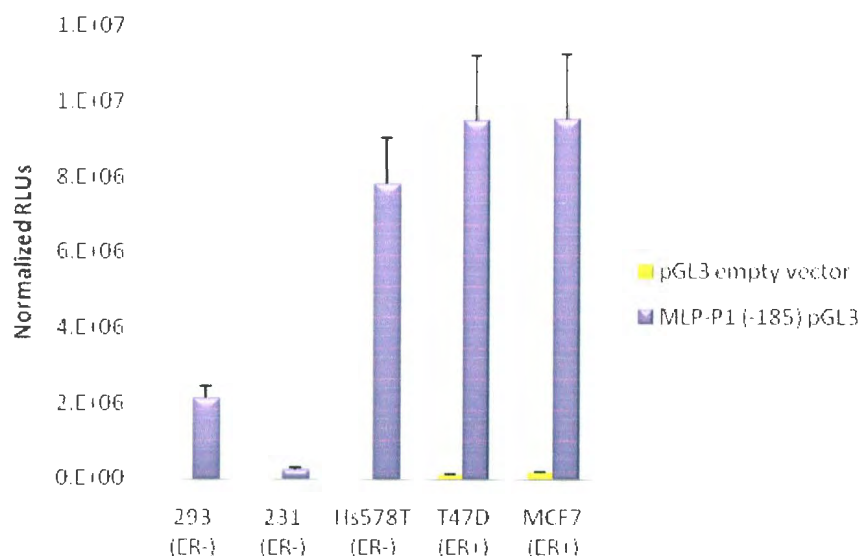
Influences that the breast cancer cellular environment have on MLP-P1 promoter activity, as well as influences of varying ER status can readily be observed following the combination and analysis of the previously performed experiments in HEK 293 cells and the multiple breast cancer cell lines<sup>17</sup>. Figure 3.6 amalgamates and organizes data from the maximum MLP-P1 promoter activity producing construct [MLP-P1 (-185) pGL3] and the pGL3 empty vector negative control data in accordance to each respective cell line. As clearly illustrated by Figure 3.6, HEK 293 cells produce much lower maximal luciferase activity levels than Hs578T (ER-) cells (with a ratio of 3.68:1 for the MLP-P1 (-185) pGL3 RLU values of Hs579T cells: HEK 293 cells), T47D (ER+) cells (with a ratio of 4.49:1 for the MLP-P1 (-185) pGL3 RLU values of T47D cells: HEK 293 cells), and MCF7 (ER+) breast cancer cells (with a ratio of 4.5:1 for the MLP-P1 (-185) pGL3 values of MCF7 cells: HEK 293 cells). Conversely, the maximal promoter activity stemming from these three latter breast cancer cell lines were very comparable to each other. Interestingly though, luciferase activity stemming from the MDA MB 231 (ER-) breast cancer cell line was much lower than that of the three aforementioned breast cancer cell lines, and was more comparable to the activity levels generated by the non-cancerous HEK 293 cells. However, MDA MB 231 cell maximal promoter activity RLU values were still

---

<sup>17</sup> All cell lines for the experiments used for this specific promoter activity comparison were transiently transfected with 0.5µg of respective MLP-P1 luciferase reporter gene deletion construct, 0.5µg of empty pCDNA3 vector (or 0.5µg of ERα pCDNA3 in the case of the HEK 293 cells), and 0.25µg of pRSVβ-gal. Cells were harvested 48 hours following transfection (52 hours following transfection for HEK 293 cells) and reporter gene assays were preformed as described in section 2.7.

much lower than those generated by HEK 293 cells (with a ratio of 8.84:1 for the MLP-P1 (-185) pGL3 RLU values of HEK 293 cells: MDA MB 231) (Fig 3.6).

The principal differences between the breast cancer cell line-specific activities did not always correlate specifically with ER status. Even though the ER+ cell line RLUs were quite proportional, the ER- Hs578T cells produced very distinct RLU levels from the ER- MDA MB 231 cells ( $p < 0.001$ ).



**Figure 3.6: Comparison of the MLP-P1 promoter activity across multiple breast cancer cell lines varying in ER status and the non-cancerous HEK 293 cell line**

MDA MB 231 cells, Hs578T cells, T47D cells and MCF7 cells were seeded at a density of  $3 \times 10^5$  cells/well and HEK 293 cells were seeded at a density of  $5 \times 10^5$  cells/well in a 6-well plate and grown in cell type specific supplemented DMEM for approximately 18 hours. Respective wells were then transfected with 0.5 $\mu$ g of MLP-P1 luciferase reporter gene deletion construct, either 0.5 $\mu$ g ER $\alpha$  pCDNA3 (for HEK 293 cells) or pCDNA3 empty vector (for all breast cancer cell lines), and 0.25 $\mu$ g of pRSV $\beta$ -gal. Cell lysates were collected 52 hours (for HEK 293 cells) or 48 hours (for all breast cancer cell lines)<sup>18</sup> following transfection, and relative luciferase units were measured. BioRad and  $\beta$ -gal assays were performed so as to collect necessary data in order to normalize the RLUs to transfection efficiency and protein levels as described in section 2.7. Each luciferase reporter gene construct was transfected into triplicate wells and the above data constitute: n=3 (293 cells), n= 4 (MB MDA 231 cells), n=3 (T47D cells), n=3 (Hs578T cells), and n=5 (MCF7 cells).

<sup>18</sup> MB MDA 231, Hs578T, T47D and MCF7 cell experimental data used for this particular comparison of MLP-P1 promoter activity was not obtained from negative control groups investigating the effect of estrogen on MLP-P1 promoter activity and therefore cells were left to grow 48 hours following initial transfection of the reporter gene deletion constructs and expression vectors.



### **3.3 *In Silico* Analysis of the MLP-P1 Promoter Proximal Region**

#### **3.3.1 Putative MLP-P1 Transcription Factor Binding Sites (TFBSs)**

This study's characterization of the MLP-P1 promoter proximal region clearly revealed that the *mier1* nucleotide sequence residing between positions -185 and -91 from the MLP-P1 TSS is involved in maximal activation at this particular promoter. In order to determine whether this specific region of MLP-P1 contains any putative human TFBSs that may be responsible for promoter activity in either a breast cancer setting specifically or within a normal physiological environment, *in silico* analysis of this region was performed. The analysis employed the following *in silico* search algorithms: transcription element search system, transcription factor binding site, gene regulation-ali baba 2.1, and the consite algorithm program as described in section 2.4.1 and 2.10. Table 3.1 lists the putative TFBSs that were specifically found within the 94bp sequence that is contained between the 5' starting end of the the maximal activation MLP-P1 reporter gene deletion construct MLP-P1 (-185) pGL3 and the 5' starting end of MLP-P1 (-91) pGL3, the luciferase reporter gene deletion construct established to contain minimal promoter activity (section 3.1.1; refer to Figure 3.7 for the full nucleotide sequence of the 94bp region between the 5' starting sites of the insertion sequences cloned into these two latter luciferase reporter gene constructs).

Moreover, as the sequences stemming from -91 to -44bp upstream of the MLP-P1 TSS were shown to be necessary for minimal activity at this promoter, this sequence was also analyzed by the four previously described transcription factor binding site *in silico* analysis programs. Table 3.3 demonstrates the putative human TFBS found in these particular regions of the MLP-P1 promoter.

```

(-249/-1528) gtggcgacca gctggggaqt ggtgcaccac cctttttttt qqccqccctct qaaqterctg

          MLP-P1(-185) pGL3
(-189/-1468) tacccccaag ctcctccgtt agcggctcgg gccgaggctc cggaatgttt gccgggcgtc

          MLP-P1(-91) pGL3
(-129/-1408) atggcgacgg tggagccctg gctcaacaag cggccgcgcg gttggctggc ggcacgaggc

```

**Figure 3.7: The sequence of the 94bp stretch of nucleotides residing between the 5' starting positions of the MLP-P1 promoter region inserts cloned into MLP-P1 (-185) pGL3 and MLP-P1 (-91) pGL3**

This figure depicts the *mier1* nucleotide sequence (Entrez Gene ID: 57708) of the MLP-P1 promoter proximal region that resides between 249 to 129 bp upstream of the putative TSS of MLP-P1. The location of the 5' starting position of the MLP-P1 promoter region insert cloned into both MLP-P1 (-185) pGL3 and MLP-P1 (-91) pGL3. As evident by this figure, the difference between these two constructs is that MLP-P1 (-185) pGL3 contains 94 extra upstream nucleotides than does MLP-P1 (-91) pGL3, the exact sequence of which is italicized and printed navy blue. It is this 94bp sequence that results in the difference of activity stemming from these two respective MLP-P1 luciferase reporter gene constructs. With reference to the nucleotide count on the left side, the first number is the nucleotide count with respect to the putative TSS of MLP-P1 where as the second number is the count with respect to the ATG translation start site.

**Table 3.1: Putative human transcription factor binding sites in the MLP-P1 promoter proximal region residing between -185 and -91bp from the MLP-P1 putative TSS**

Transcription Element Search System	Transcription Factor Binding Site <sup>19</sup>	Gene Regulation- Ali Baba 2.1	CONSITE
c-myb	SRY	<b><i>Sp1</i></b>	NRF-2
AP-2α A	HSF2		TEF-1
<b><i>Sp1</i></b>	<b><i>CREB</i></b>		FREAC-4
<b><i>CREB</i></b>	ZID		E2F
AP-1	AP-1		<b><i>CREB</i></b>
E2F + p107			
CAC-Binding Protein			
c-Myc			
GCF			

- Human TFBSs found by at least two different search engine/algorithm programs that were located in the same position of the MLP-P1 promoter region encompassing the same nucleotides are bolded, italicized, and outlined by a red box.

<sup>19</sup> This particular search engine was employed using 85% & 80% homology search parameters.

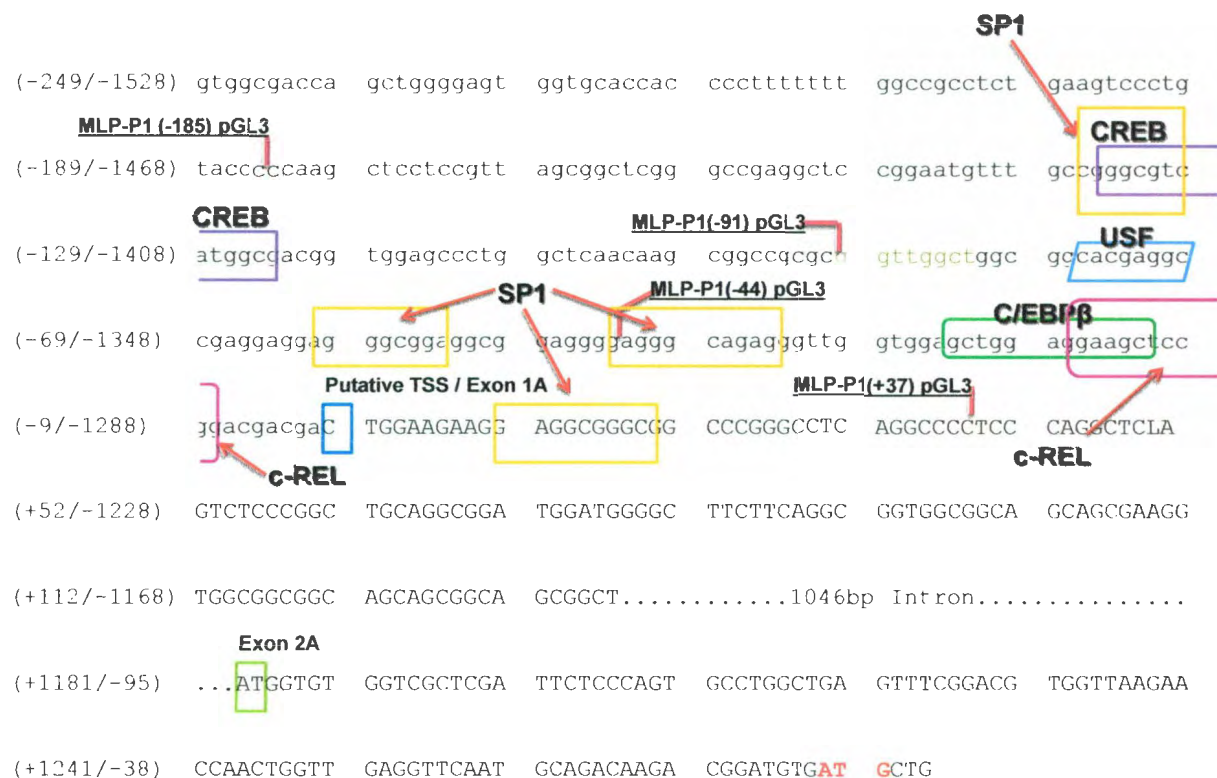


**Table 3.2: Putative human transcription factor binding sites in the MLP-P1 proximal promoter region outside residing between -91bp to -44bp from the MLP-P1 putative TSS**

Transcription Element Search System	Transcription Factor Binding Site	Gene Regulation- Ali Baba 2.1	CONSITE
C/EBP $\alpha$	<b><i>USF</i></b>	<b><i>Sp1</i></b>	CREB
ETF	MZF1	YY1	<b><i>USF</i></b>
<b><i>Sp1</i></b>	<b><i>C/EBP<math>\beta</math></i></b>	c-Ets-1	MAX
AP-2 $\alpha$ A	<b><i>c-REL</i></b>	C/EBP $\alpha$	p65
AP-2 $\alpha$ B	RELA NF-Kappa		<b><i>c-REL</i></b>
Maz	HSF2		SAP-1
LF-A1	<b><i>Sp1</i></b>		Thing1-E47
CAC-Binding Protein			
<b><i>C/EBP<math>\beta</math></i></b>			
C/EBP $\delta$			
c-Ets-2			
TCF-2 $\alpha$			
PEA3			
GCF			

- Human TFBSs found by at least two different search engine/algorithm programs that were located in the same position of the MLP-P1 promoter region encompassing the same nucleotides are bolded, italicized, and outlined by a red box.

The exact sequences of the TFBSs listed in Tables 3.1 & 3.2 that were annotated with red boxes are depicted in Figure 3.8, which plots the TFBSs with reference to their respective location within the MLP-P1 promoter proximal region. These particular human TFBSs were detected by at least two different *in silico* sequence analysis programs.



**Figure 3.8: Putative human transcription factor binding sites in the *mier1* MLP-P1 promoter proximal region**

This figure illustrates the location and sequences of the following putative human TFBSs: Sp1, CREB, USF, C/EBP $\beta$ , and C-REL. These putative TFBS were found and described in at least two of the employed transcription factor *in silico* analysis programs as described in section 2.4.1 and 2.10. The locations of the 5' starting positions of the insertion sequence of each MLP-P1 promoter proximal luciferase reporter gene deletion construct are also indicated. The first number on the left side is the nucleotide count with respect to the putative TSS of MLP-P1 where as the second number is the count with respect to the MLP-P1 ATG translation start site.



### 3.3.2 Analysis of the Percent Homology of the MLP-P1 Promoter Proximal Region Sequence Conservation Across Various Species

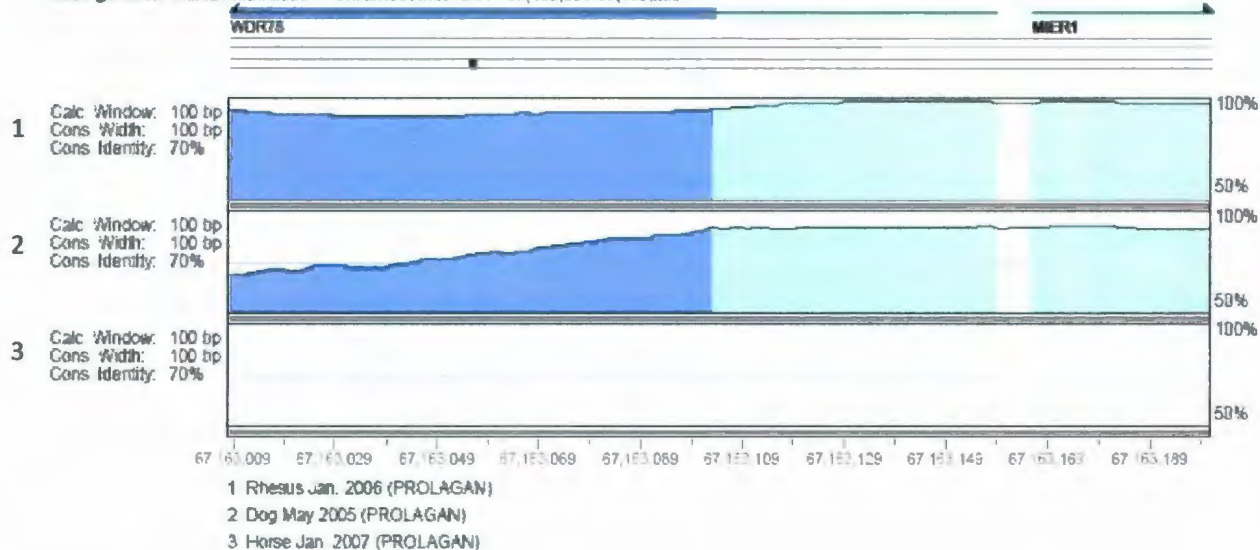
The analysis of evolutionarily conserved sequences across multiple species allows scientists to determine the functional and/ or regulatory importance of specific genetic sequences (Dubchak *et al.*, 2009). Sequence conservation highlights genetic segments or elements likely to mediate important biological function, as it is assumed that functionally important elements evolve more slowly than non-functional genomic regions due to selective constraints (Shah *et al.*, 2004). In order to further stress the functional importance of the MLP-P1 sequences found to be involved in eliciting maximal and minimal promoter activity, this study analyzed the amount of sequence conservation that these particular regions of *mier1* portray across various species using the VISTA portal system computational tools as described in section 2.10. The species whose sequences were compared to the human MLP-P1 promoter region were as follows: rhesus (chimpanzee), mouse, rat, dog, horse, and chicken.

Table 3.3 gives the exact percent sequence conservation between the denoted human MLP-P1 promoter sequences when compared to the analyzed species. It is important to note that the location on chromosome 1 whereby the maximal promoter activity region of MLP-P1 (between -185 to -91bp from the MLP-P1 putative TSS) is chr 1: 67,163,045-67,163,139. The location on chromosome 1 for the minimal promoter activity region of MLP-P1 (between -91 to -44bp from the MLP-P1 putative TSS) is chr 1: 67,163,139- 67, 163, 186. Figure 3.9 furthermore illustrates a graphical depiction of the sequence homology between these various species that was outlined in Table 3.3. Refer to Appendix 7 for the exact sequence nucleotides of each species sequence aligned to that of human MLP-P1 promoter sequences.

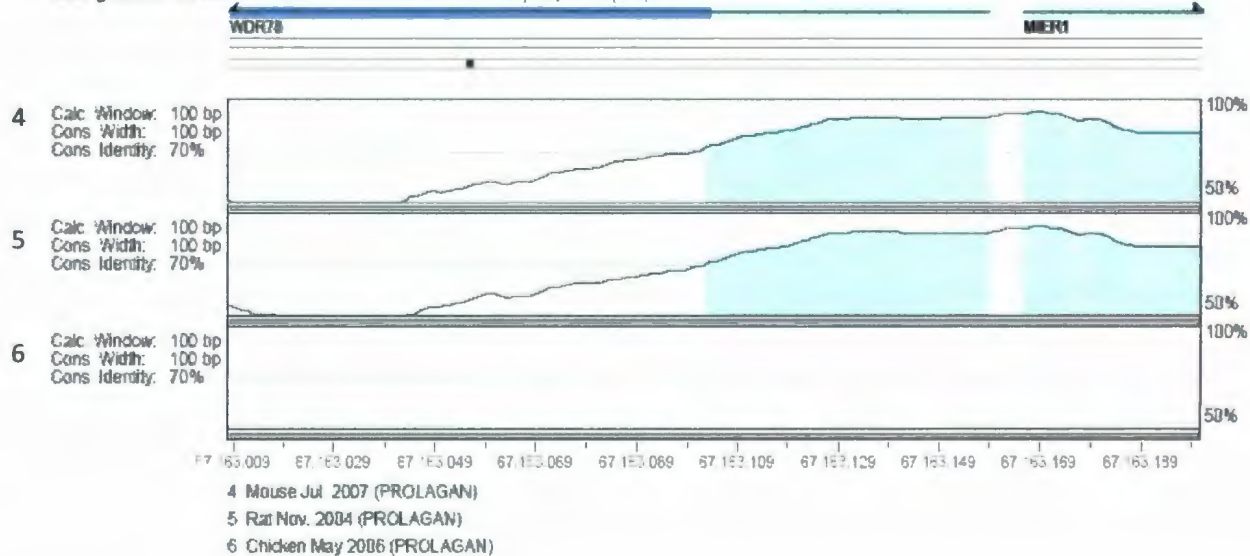
**Table 3.3: Percent conservation of MLP-P1 maximal and minimal promoter regions across six species compared to human sequences**

<b>Species Compared</b>	<b>Region of Conservation MLP-P1 promoter with reference to location on Human Chromosome 1</b>	<b>Region of Conservation of MLP-P1 with reference to location on compared species' chromosome</b>	<b>% Conservation</b>	<b>Total % Conservation for area chr1: 67, 163, 045-67, 163, 200</b>
<b>Human vs. Rhesus</b>	67,163,008-67,163,102	chr1: 69,702,667-69,702,760	91.6	95.2
	67,163,103-67,163,158	chr1: 69,702,761-69,702,816	98.2	
	67,163,166-67,163,200	chr1: 69,702,824-69,702,858	100	
<b>Human vs. Mouse</b>	67,163,103-67,163,158	chr4: 102,786,872-102,786,926	89.3	92.3
	67,163,166-67,163,200	chr4: 102,786,934-102,786,968	97.1	
<b>Human vs. Rat</b>	67,163,103-67,163,158	chr5: 124,048,297-124,048,351	87.5	91.2
	67,163,166-67,163,200	chr5: 124,048,359-124,048,393	97.1	
<b>Human vs. Dog</b>	67,163,008-67,163,102	chr5: 46,592,439-46,592,347	78.9	86.0
	67,163,103-67,163,158	chr5: 46,592,346-46,592,291	91.1	
	67,163,166-67,163,200	chr5: 46,592,284-46,592,250	97.1	
<b>Human vs. Horse</b>	n/a	chr5: n/a	0	0
<b>Human vs. Chicken</b>	n/a	chr8: n/a	0	0

Base genome: Human Mar. 2006 Chromosome: chr1 67,163,008-67,163,200



Base genome: Human Mar. 2006 Chromosome: chr1 67,163,008-67,163,200



**Figure 3.9: Graphical depiction of sequence homology of *mier1*'s MLP-P1 promoter region across six different species compared to the human sequence**

Annotations:  
 ▲ Gene    ■ JTR  
 ■ Exon

Human *mier1* MLP-P1 promoter sequences consisting of the established minimal and maximal promoter regions were compared to the following species: rhesus (chimpanzee), dog, horse, mouse, and chicken. This comparison, using the vista browser tool, aligned pre-computed and established genome sequences in order to calculate the percent conservation between these respective species as compared to the human sequence. The percent conservation is depicted on the y axis of the graphs in this figure, which was obtained following the vista browser tool analysis. Further annotations appearing in this figure are explained in the legend located to the left.



### 3.4 Comparison of MLP-P1 vs MAEP-P2 Promoter Activity in Breast Cancer Cells

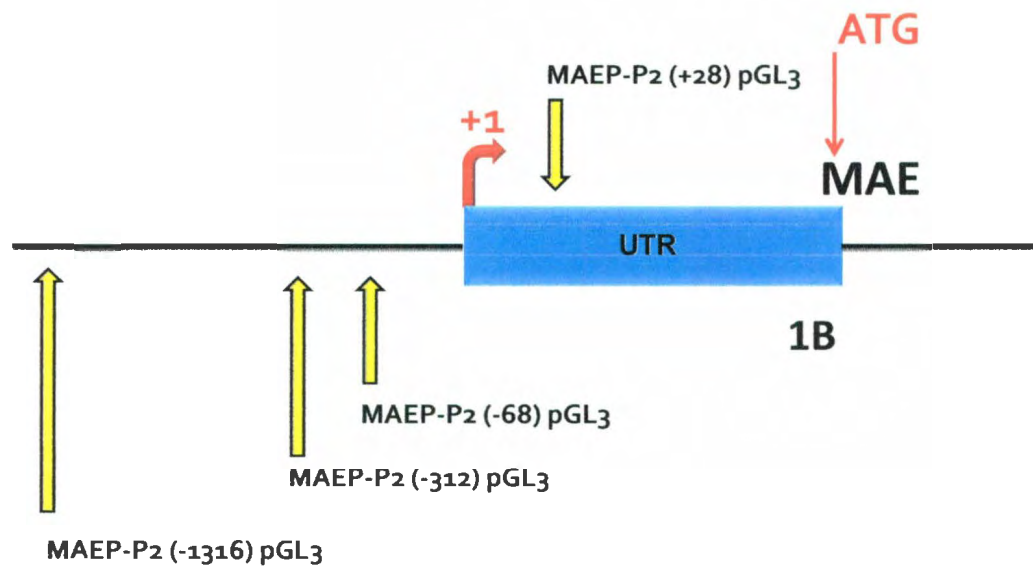
As previously discussed in section 1.3.1.2, alternate promoter usage is implicated in the generation of tissue specific expression of various gene protein isoforms and highly contributes to the diversity of a given proteome. Additionally there is growing evidence linking aberrant use of multiple promoters to cancer initiation and progression, as several oncogenes and tumour suppressor genes have multiple promoters. Moreover, the aberrant use of one promoter over another for various genes has been shown to be directly linked to cancerous cell growth (Agarwal *et al.*, 1996; Arce *et al.*, 2006; Davuluri *et al.*, 2008). As *mier1* contains two alternate promoters directing the transcription of its various protein isoforms, this gene can potentially be differentially regulated through aberrant promoter usage. Moreover, as briefly discussed in section 1.5.6 and further analyzed in section 3.5, there are significant possible implications to the preferential usage of MLP-P1, such as: an abnormal increase in transcription at MLP-P1 may result in the generation of a large proportion of proteins harbouring putative NESs and/ or putative transmembrane domains. As such a phenomenon may help explain some of the molecular mechanisms contributing to the abnormal change in subcellular localization of MIER1 $\alpha$  from the nucleus to the cytoplasm when breast cells progress to a more invasive state, it becomes even more pertinent to investigate and compare activation at both *mier1* promoters in breast cancer.

### 3.4.1 MAEP-P2 Promoter Activity Pattern in Breast Cancer Cells

Many factors can induce alternate promoter usage, one of the principle governing components being triggers from the cellular environment; such as: tissue-specific TFs, coregulators, and even abnormal regulatory factors present throughout chronic disease, such as cancer. In order to initially investigate whether or not either of the *mier1* promoters are differentially regulated in breast cancer cells, the luciferase reporter gene deletion construct approach used throughout this study was employed. As reported in section 1.5.2.1, the MAEP-P2 promoter had already been characterized by our laboratory prior to this study (Ding *et al.*, 2004). However, this characterization did not look at multiple sections of the MAEP-P2 promoter in various breast cancer cell lines. Therefore, this study's first step was to investigate MAEP-P2 promoter activity in breast cancer cells by initially transfecting MCF7 cells with four separate MAEP-P2 reporter gene deletion constructs that incorporated MAEP-P2 promoter proximal as well as promoter distal regions spanning up until 1316 bp upstream of the TSS (-1464 upstream of the MAEP-P2 ATG translation start site).

The constructs used for this initial characterization were as follows: MAEP-P2 (-1316) pGL3, MAEP-P2 (-312) pGL3, MAEP-P2 (-68) pGL3, and MAEP-P2 (+28) pGL3 (Fig 3.10). For further details with regards to the exact MAEP-P2 promoter region nucleotide sequences inserted into these luciferase reporter gene deletion constructs refer to Figure 2.4 and Appendix 3. The pGL3 empty vector was used as a negative control for background luciferase activity. Forty-eight hours following transfection, cells were harvested and luciferase,  $\beta$ -gal, and bio-rad reporter assays were performed as described previously (section 2.7). Figure 3.11 depicts results of

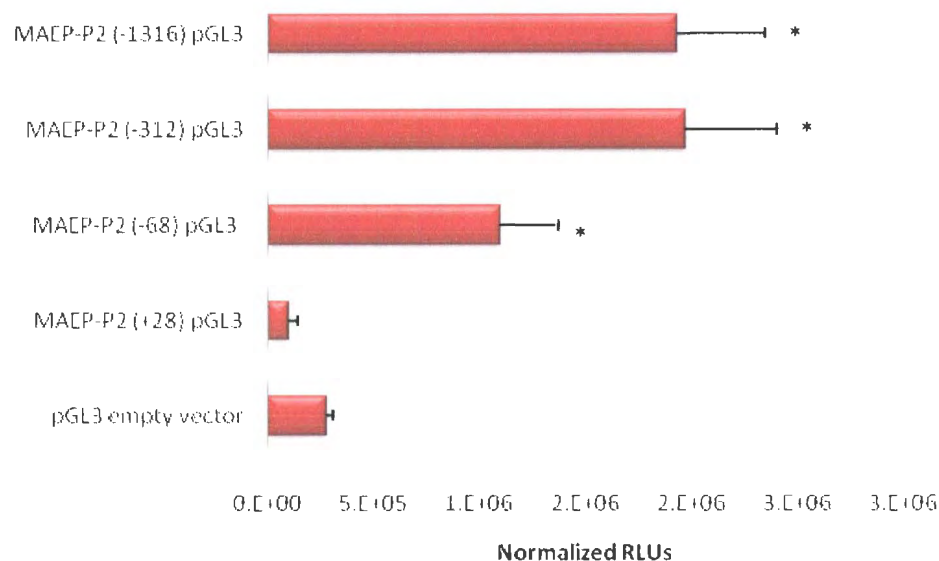
luciferase activity measured, whereby the MAEP-P2 (-68) pGL3 reporter gene deletion construct is the smallest sequence significantly up-regulated above the pGL3 empty vector control ( $p < 0.001$ ). Furthermore, Figure 3.11 demonstrates that MAEP-P2 (-312) pGL3 contained the MAEP-P2 nucleotide sequence that produced maximal promoter activity in MCF7 cells. Refer to Appendix 5 for results following these exact same experiments in HEK 293 cells, which further recapitulate this established promoter activity pattern.



**Figure 3.10: Schematic of luciferase reporter gene deletion constructs used for the initial characterization of MAEP-P2 promoter activity in MCF7 cells**

This schematic depicts the MAEP-P2 promoter region and the 5' starting position of the MAEP-P2 putative promoter region inserts cloned into each individual MAEP-P2 luciferase reporter gene deletion construct that was used in this current section (section 3.4.1). Additional notation depicted in this schematic describes the MAEP-P2 proximal promoter region which is further described in section 2.2.3.2.



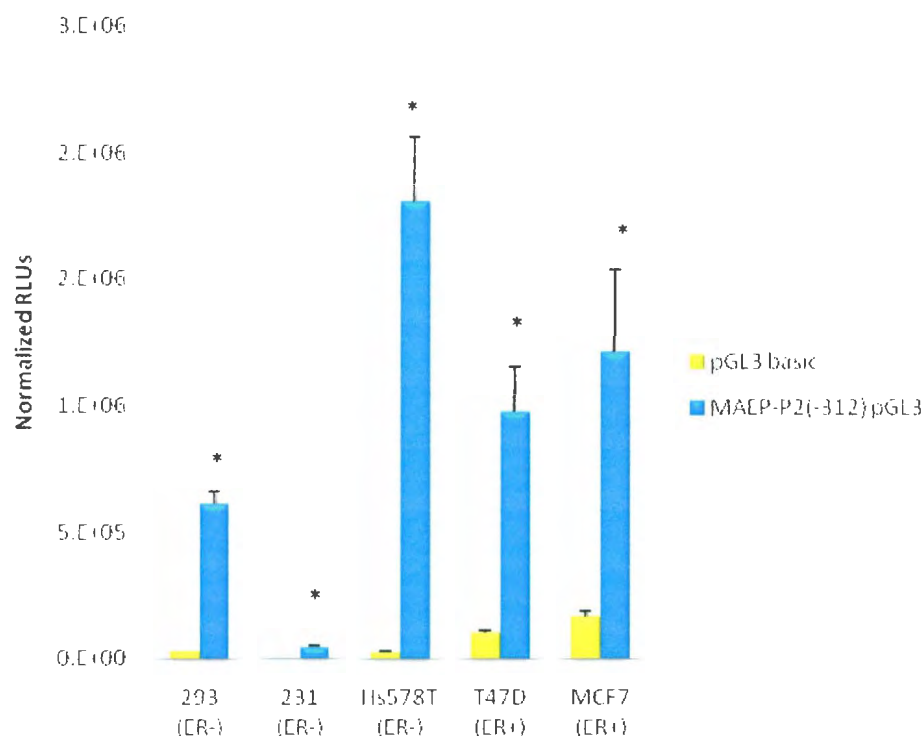


**Figure 3.11: Characterization of the MAEP-P2 promoter region of *miel1* in MCF7 breast cancer cells**

MCF7 cells were seeded at a density of  $3 \times 10^5$  cells/well of a 6-well plate and grown in DMEM for approximately 18 hours. Cells were then transfected with 1  $\mu$ g of MLP-P1 luciferase reporter gene deletion construct and 0.25  $\mu$ g of pRSV $\beta$ -gal. Cell lysates were collected 48 hours following transfection and relative luciferase units were measured. BioRad and  $\beta$ -gal assays were performed so as to collect necessary data in order to normalize the RLUs to protein levels and transfection efficiency section 2.8. Each luciferase reporter gene construct was transfected into triplicate wells and the above data constitute an n=5. Asterisks denote statistically significant differences between the RLUs stemming from the luciferase reporter gene deletion constructs and the pGL3 empty vector negative control ( $p < 0.001$  for each comparison).

Similar to the investigation of MLP-P1 promoter activity described in section 3.2.3, MAEP-P2 promoter activity was analyzed across multiple breast cancer cell lines in order to: 1) initially investigate whether ER status could potentially affect activation at MAEP-P2 in breast cancer cells, and 2) investigate the MAEP-P2 promoter activity patterns in breast cancer cells and to compare these results to a non-cancerous cell line (HEK 293 cells). The procedure briefly outlined in section 3.2.3 and extensively explained in sections 2.5-2.7 was employed for these experiments, using the pGL3 empty vector as the negative control and the maximum MAEP-P2 promoter activity producing construct, MAEP-P2 (-312) pGL3, in order to monitor MAEP-P2 promoter activity. Figure 3.12 depicts and compares results of these independent experiments across the HEK 293, MDA MB 231, T47D, Hs578T, and MCF7 cell lines. Luciferase data demonstrates that MAEP-P2 was activated at different levels in these cell lines, with the ER- MDA MB 231 breast carcinoma cell line displaying the lowest level in relative luciferase activity. All MAEP-P2 (-312) pGL3 constructs in every cell line produced significantly higher promoter activity than the pGL3 empty vector negative control ( $p < 0.001$ ), which demonstrates that there is activity at this promoter in each cell line examined. This data also paralleled the observed pattern across cell line-specific MLP-P1 promoter activity whereby MDA MB 231 and HEK 293 cells also generated lower amounts of promoter activity levels at the MAEP-P2 promoter than the T47D, MCF7, and Hs578T cells. Contrary to the MLP-P1 promoter though, there was not as a prominent difference between the promoter activity levels of HEK 293 cells when compared to T47D, MCF7, and Hs578T at the MAEP-P2 promoter (Fig 3.12). Furthermore, the differences between cell line-specific activities did not always consistently correlate with ER status, as the

ER- Hs578T cells produced very distinct RLU levels from the ER- MDA MB 231 cells. The ER+ cell line RLU levels; however, were comparable.



**Figure 3.12: Analysis of the MAEP-P2 promoter activity across multiple breast cancer cell lines varying in ER Status and the HEK 293 non-cancerous cell line**

MB MDA 231 cells, Hs578T cells, T47D cells and MCF7 cells were seeded at a density of  $3 \times 10^5$  cells/well and HEK 293 cells were seeded at a density of  $5 \times 10^5$  cells/ well in a 6-well plate and grown in cell type specific supplemented DMEM for approximately 18 hours. Respective wells were then transfected with 0.5 $\mu$ g of MAEP-P2 luciferase reporter gene deletion construct, either 0.5 $\mu$ g ER $\alpha$  pCDNA3 (for HEK 293 cells) or pCDNA3 empty vector (for all breast cancer cell lines), and 0.25 $\mu$ g of pRSV $\beta$ -gal. Cell lysates were collected 52 hours (for HEK 293 cells) or 48 hours (for all breast cancer cell lines) following transfection, and relative luciferase units were measured. BioRad and  $\beta$ -gal assays were performed so as to collect necessary data in order to normalize the RLUs to protein levels as well as transfection efficiency levels as described in section 2.8. Each luciferase reporter gene construct was transfected into triplicate wells and the above data constitute: n=3 (293 cells), n= 4 (MB MDA 231 cells), n=3 (T47D cells), n=3 (Hs578T cells), and n=5 (MCF7 cells). Asterisks denote statistically significant difference between the MAEP-P2 (-312) pGL3 luciferase reporter gene deletion construct and the pGL3 empty vector for each respective cell line ( $p < 0.001$ ).

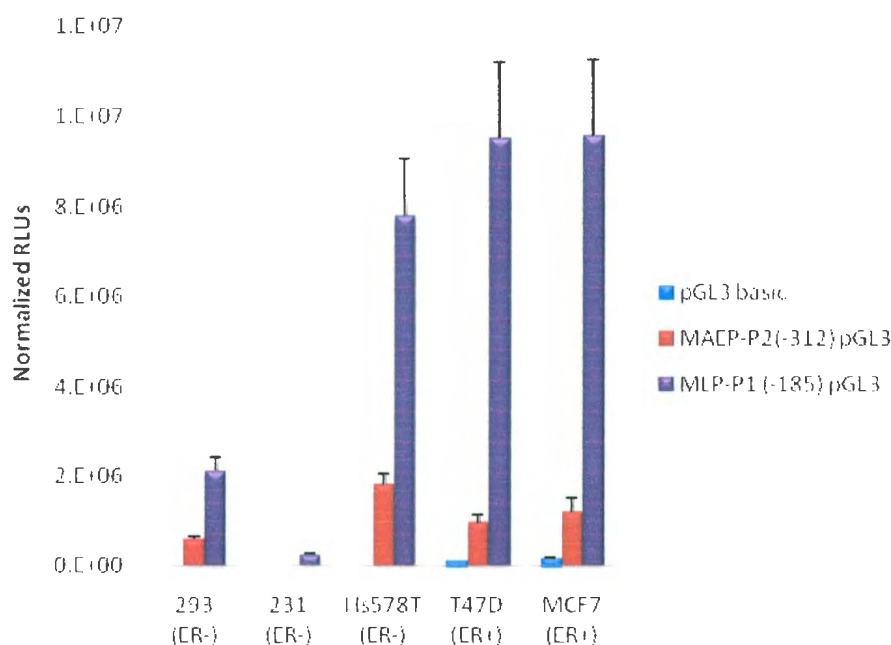


### 3.4.2 MLP-P1 vs MAEP-P2 Promoter Activity in Breast Cancer Cells

In order to: 1) determine whether one of *mier1*'s promoters is increasingly activated over the other in breast cancer cells compared to a non-neoplastic cells, and 2) investigate whether the ER status of the breast cancer cell lines further affects the degree of promoter activity stemming from one *mier1* promoter over the other, luciferase data following previously described transient transfections of four breast cancer cell lines and the non-cancerous HEK 293 cell line was compared. These experiments employed both the MAEP-P2 and MLP-P1 luciferase reporter gene deletion constructs used in sections 3.2.3 and 3.3.2 [MLP-P1 (-185) pGL3, MLP-P1 (-91) pGL3, MLP-P1 (-44) pGL3, and MLP-P1 (+37) pGL3 for MLP-P1 and MAEP-P2 (-312) pGL3 and MAEP-P2 (+28) pGL3 for MAEP-P2], as well as the pGL3 empty vector negative control. Along with these *mier1* luciferase reporter gene deletion constructs, the pRSV- $\beta$  gal expression vector, and either the human ER $\alpha$  pCDNA3 (for HEK 293 cells) or pCDNA3 empty vector (all breast cancer cell lines) were also transfected into each well as described previously (section 2.5). For ease of comparison, Figure 3.13 displays the respective luciferase activities stemming from only the maximal promoter activity-producing luciferase reporter gene deletion constructs for each *mier1* promoter [MLP-P1 (-185) pGL3 and MAEP-P2 (-312) pGL3]<sup>20</sup>. This data precisely demonstrates that the MLP-P1 promoter produced significantly higher levels of luciferase activity above the MAEP-P2 promoter in each cell line tested ( $p < 0.001$ ) (Fig 3.13). The extent by which MLP-P1 was increasingly

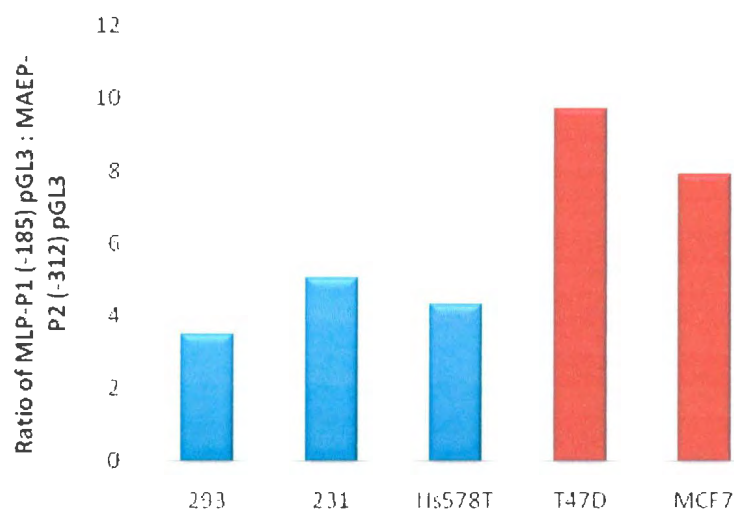
<sup>20</sup> Data collected from the other MLP-P1 and MAEP-P2 luciferase reporter gene deletion constructs (MLP-P1 (-91) pGL3, MLP-P1 (-44) pGL3, and MLP-P1 (+37) pGL3 for the MLP-P1 promoter, and MAEP-P1 (+28) pGL3 for the MAEP-P2 promoter) are illustrated in figure 3.4 for the MLP-P1 promoter and in appendix 6 for the MAEP-P2 promoter.

activated over MAEP-P2 did differ between the respective cell lines, and is succinctly illustrated by Figure 3.13 and Figure 3.14. Figure 3.14 specifically depicts the respective fold increases between these two constructs that generate maximum promoter activity at both *mier1* promoters. Interestingly, the degree to which MLP-P1 was stimulated over MAEP-P2 was highest in ER+ cell lines.



**Figure 3.13: Comparison of MLP-P1 vs. MAEP-P2 promoter activity across multiple breast cancer cell lines varying in ER status and the non-cancerous HEK 293 cell line**

MB MDA 231 cells, Hs578T cells, T47D cells and MCF7 cells were seeded at a density of  $3 \times 10^5$  cells/well and HEK 293 cells were seeded at a density of  $5 \times 10^5$  cells/well in a 6-well plate and grown in cell type specific supplemented DMEM for approximately 18 hours. Respective wells were then transfected with either 0.5 $\mu$ g of MAEP-P2 luciferase reporter gene deletion construct or MLP-P1 luciferase reporter gene deletion constructs, and either 0.5 $\mu$ g ER $\alpha$  pCDNA3 (for HEK 293 cells) or pCDNA3 empty vector (for all breast cancer cell lines), and 0.25 $\mu$ g of pRSV $\beta$ -gal. Cell lysates were collected 52 hours (for HEK 293 cells) or 48 hours (for all breast cancer cell lines) following transfection, and relative luciferase units were measured. BioRad and  $\beta$ -gal assays were performed so as to collect necessary data in order to normalize the RLUs to transfection efficiency and protein levels as described in section 2.7. Each luciferase reporter gene construct was transfected into triplicate wells and the above data constitute: n=3 (293 cells), n= 4 (MB MDA 231 cells), n=3 (T47D cells), n=3 (Hs578T cells), and n=5 (MCF7 cells).



**Figure 3.14: MLP-P1 (-185) pGL3 vs MAEP-P2 (-312) pGL3 luciferase activity comparison across various breast cancer cell lines and HEK 293 cells**

MB MDA 231 cells, Hs578T cells, T47D cells and MCF7 cells were seeded at a density of  $3 \times 10^5$  cells/well and HEK 293 cells were seeded at a density of  $5 \times 10^5$  cells/well in a 6-well plate and grown in cell type specific supplemented DMEM for approximately 18 hours. Respective wells were then transfected with either 0.5 $\mu$ g of MAEP-P2 luciferase reporter gene deletion construct or MLP-P1 luciferase reporter gene deletion constructs, and either 0.5 $\mu$ g ER $\alpha$  pCDNA3 (for HEK 293 cells) or pCDNA3 empty vector (for all breast cancer cell lines), and 0.25 $\mu$ g of pRSV $\beta$ -gal. Cell lysates were collected 52 hours (for HEK 293 cells) or 48 hours (for all breast cancer cell lines) following transfection, and relative luciferase units were measured. BioRad and  $\beta$ -gal assays were performed so as to collect necessary data in order to normalize the RLUs to transfection efficiency and protein levels as described in section 2.7. Each luciferase reporter gene construct was transfected into triplicate wells and the above data constitute:  $n=3$  (293 cells),  $n=4$  (MB MDA 231 cells),  $n=3$  (T47D cells),  $n=3$  (Hs578T cells), and  $n=5$  (MCF7 cells). This figure graphs the ratio of the normalized relative luciferase units stemming from the maximum promoter activity producing miel promoter luciferase reporter gene deletion construct for each cell line. The ratio is as follows: MLP-P1 (-185) pGL3: MAEP-P2 (-312) pGL3 and is specified by the y axis. The cell line is specified on the x axis. ER- cell lines are represented by the blue columns while the ER+ cell lines are represented by the red columns.



### **3.5 Effect of Incorporation of Exon 3A on MIER1 $\alpha$ Subcellular Localization**

Section 1.5.6 briefly introduced a potential implication to preferential usage of MLP-P1, being the inclusion of *mier1* alternate exon 3A which may encode a NES and/ or transmembrane domain. Section 3.4 revealed MLP-P1 as eliciting higher levels of promoter activity than MAEP-P2 in HEK 293 cells and in each breast carcinoma cell line tested, an occurrence that was more pronounced in ER+ breast cancer cell lines. Due to such predominant promoter activity production at MLP-P1 in these cell lines, and because an increased inclusion of exon 3A could serve as a possible mechanism explaining MIER1 $\alpha$ 's aberrant subcellular localization in invasive breast cancer, this study included as a final objective to investigate the nature of the putative NES and/ or transmembrane domain of exon 3A and to observe whether it can perpetuate a change in subcellular localization from the nucleus to the cytoplasm and/ or plasma membrane.

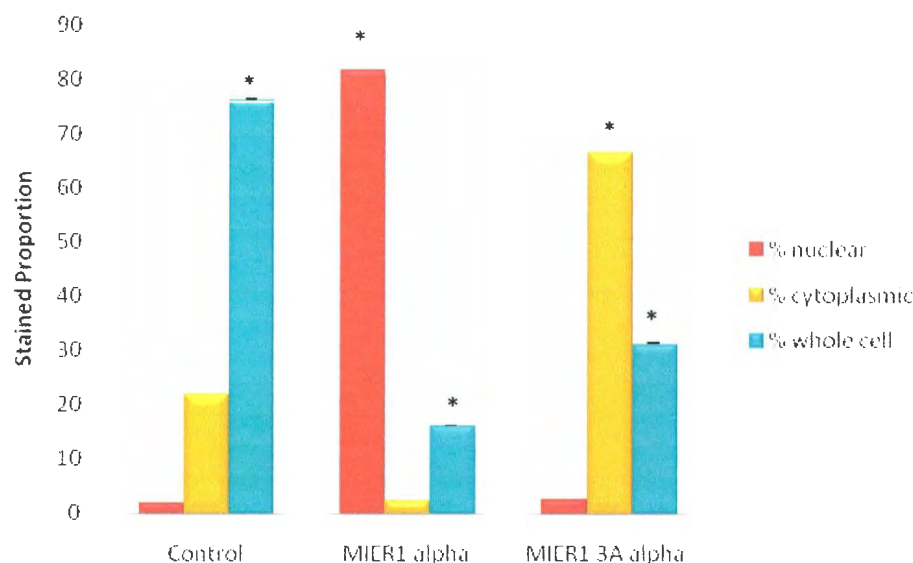
#### **3.5.1 Subcellular Localization of *mier1* Isoforms derived from Transcripts either containing or lacking the 74bp Insert encoded by Exon 3A in MCF7 Breast Carcinoma Cells**

Transient transfections of expression vectors encoding MYC-tagged MIER1 $\alpha$  with or without the 74bp insert encoded by exon 3A (pCS3+MT *mier1* $\alpha$  and pCS3+MT *mier1* $\alpha$  exon 3A respectively), as well as a MYC-tagged control vector (pCS3+MT) were initially performed in MCF7 cells grown in 8-well chamber slides as described in section 2.8. Forty-eight hours following transfection, ICC was performed using antibodies specific for the MYC tag present in the specific proteins

produced by the transfected expression vectors. Stained cells were counted and the location of the staining, whether nuclear, cytoplasmic, or whole cell staining was noted. Specifically, staining was considered nuclear if only the nucleus was exclusively stained. Consequently, the cytoplasmic category consisted of cells that exclusively displayed cytoplasmic staining, while the cells were considered to exhibit whole cell staining when both nuclear and cytoplasmic staining appeared in the cell at approximately equal levels. The histogram of Figure 3.15 represents the respective proportion of cells that were observed in each category (ie: cytoplasmic, nuclear, or whole cell) for each *mier1* isoform and the tag control<sup>21</sup>. Cells transfected with pCS3+MT *mier1α* (denoted MIER1 alpha in Fig 3.15) mostly resulted in nuclear staining (82%) followed by some whole cell staining (16%) and virtually no cytoplasmic only staining (2%). The MCF7 cells transfected with pCS3+MT *mier1α* exon 3A (denoted MIER1 3A alpha in Fig 3.15), on the other hand, mostly displayed cytoplasmic staining (66%), with approximately 31% whole cell staining and virtually no nuclear only staining (3%). The MYC-tagged control displayed majority whole cell staining (76%), 22% cytoplasmic only staining and minimal nuclear only staining (2%). Representative pictures taken of these particular transfected MCF7 cells, as well as non-transfected controls stained or not stained with primary antibody are illustrated in Figure 3.16. Note that the non-transfected controls did not stain (Fig 3.16 M-P).

---

<sup>21</sup> The tag control is simply the MYC tag-containing pCS3 vector: pCS3+MT

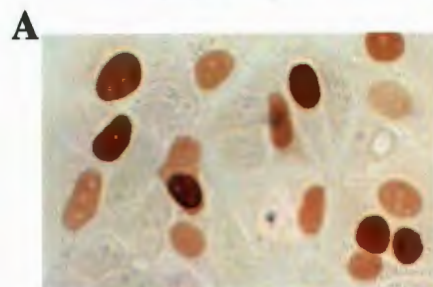


**Figure 3.15: Subcellular localization data of *mier1* isoforms derived from transcripts either containing or lacking the 74bp insert encoded by exon 3A in MCF7 breast carcinoma cells**

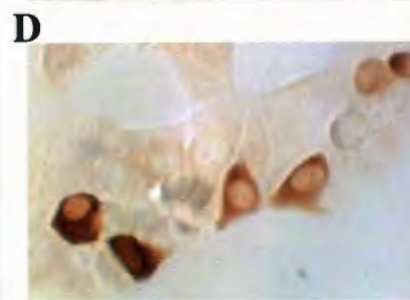
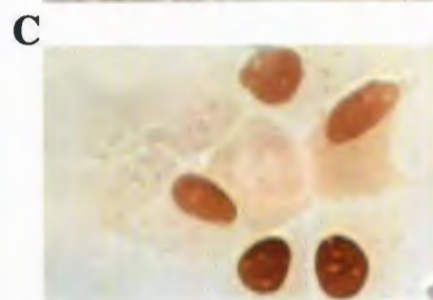
MCF7 cells were seeded at a density of 15,000 cells/ well in an 8-well chamber slide and grown in MCF7 specific supplemented DMEM for approximately 18 hours. Chamber slide wells were then transfected with 0.26µg of either pCS3+MT*mier1a*, pCS3+MT *mier1a* exon 3A , the MYC-tagged control vector (pCS3+MT), or were left untransfected. Forty eight hours following transfection, cells were fixed with 4% paraformaldehyde and ICC was performed with an antibody specific for the MYC tag. Stained cells were counted and categorized as either having nuclear, cytoplasmic, or whole cell staining. The y axis represents the proportion of cells in each category whereby the x axis denotes the specific transfected *mier1* isoforms and the tagged control whereby the isoform produced from transfection with pCS3+MT*mier1a* is termed MIER1 alpha and the isoform produced from produced from pCS3+MT *mier1a* exon 3A transfection is termed MIER1 3A alpha. "Control" represents the MYC tagged-control pCS3+MT. n=4 for both MIER1 alpha and MIER1 3A alpha while n=2 for the control tag. Astrixes represent significant differences between each category: MIER1 alpha: nuclear staining vs cytoplasmic (p<0.001), and nuclear vs whole cell staining (p<0.001); MIER1 3A alpha: nuclear staining vs cytoplasmic (p<0.001), nuclear staining vs whole cell staining (p<0.001); tag control: nuclear vs whole cell staining (p<0.001), and cytoplasmic vs whole cell staining (p<0.001). Comparison between MIER1 constructs: 1. MIER1 alpha nuclear staining vs MIER1 3A alpha nuclear stainin= p<0.001; 2. MIER1 alpha cytoplasmic staining vs MIER1 3A alpha cytoplasmic staining= p<0.001; 3. MIER1 alpha whole cell staining vs MIER1 3A alpha whole cell staining= p>0.05.



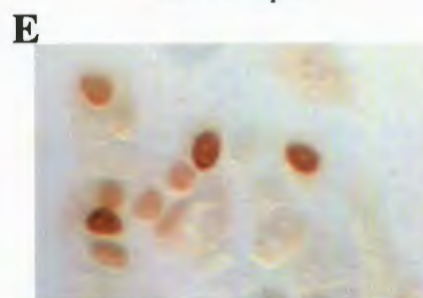
*MIER1 alpha*



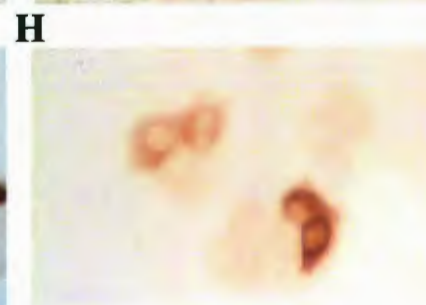
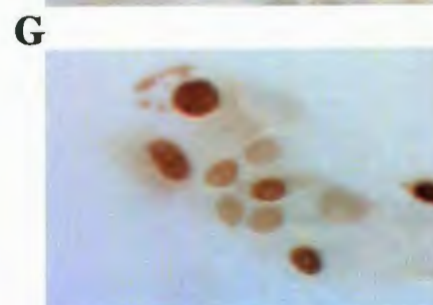
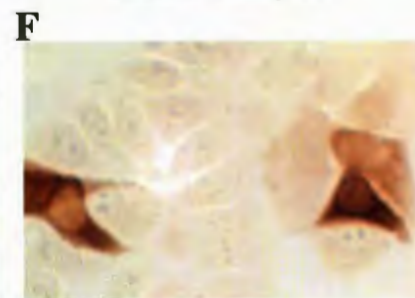
*MIER1 3A alpha*



*MIER1 alpha*



*MIER1 3A alpha*





**Figure 3.16: Subcellular localization pictures of *mier1* isoforms derived from transcripts either containing or lacking the 74bp insert encoded by exon 3A in MCF7 breast carcinoma cells**

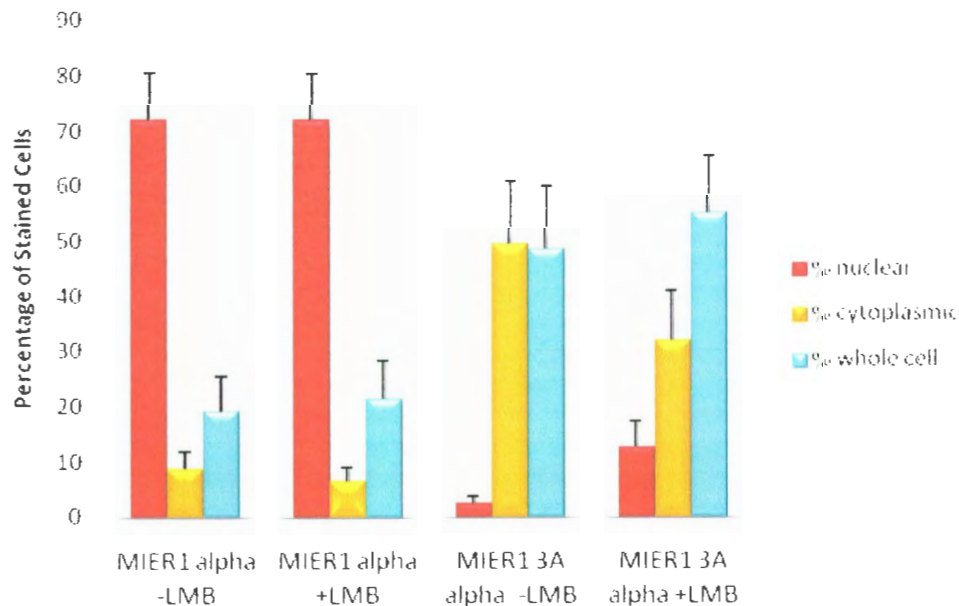
MCF7 cells were seeded at a density of 15,000 cells/ well in an 8-well chamber slide and grown in MCF7 specific supplemented DMEM for approximately 18 h. Chamber slide wells were then transfected with 0.26µg of either pCS3+MT*mier1a*, pCS3+MT *mier1a* exon 3A , the MYC-tagged control vector (pCS3+MT), or were left untransfected. Forty eight hours following transfection, cells were fixed with 4% paraformaldehyde and ICC was performed with an antibody specific for the MYC tag. Stained cells were counted and categorized as either having nuclear, cytoplasmic, or whole cell staining. Fig 3.16 A, C, E, & G represent different cells transfected with pCS3+MT*mier1a* (denoted MIER1 alpha), Fig 3.16 B, D, F, and H represent different cells transfected with pCS3+MT *mier1a* exon 3A (denoted MIER1 3A alpha). All these latter pictures were taken in bright field. Fig 3.16 I-L represent the MYC-tagged control vector while Fig 3.16 M-P represent additional controls. Specifically, Fig 3.16 M & N consist of cells that were not transfected with any vector expression vector and that were not stained with a primary antibody (M= bright field, N=phase contrast), while Fig 3.16 O & P consist of cells that were not transfected but that were however stained with the anti-MYC primary antibody as described in section 2.8 (O= bright field, P=phase contrast).

### **3.5.2 Subcellular Localization of *mier1* Isoforms derived from Transcripts either containing or lacking the 74bp Insert encoded by Exon 3A following Leptomycin B Treatment in MCF7 Cells**

Nucleocytoplasmic trafficking, and thus the import and export of various molecules and proteins to and from the nucleus, is mediated by nuclear transport receptors belonging to the importin family (La Cour *et al.* 2003). In the case of nuclear exportation specifically, proteins harbouring leucine-rich nuclear export signals (NESs) are translocated to the cytoplasm primarily through the CRM1-dependent pathway (Kosugi *et al.*, 2008; La Cour *et al.* 2003). CRM1, which is a major cellular exportin, can be inhibited by the potent antifungal antibiotic leptomycin B (LMB) through direct alkylation of CRM1 cystine 525 (Dong *et al.* 2009, Kudo *et al.* 1998; Kudo *et al.*, 1999). This function of LMB has effectively allowed LMB treatment to serve as a means to verify whether or not subcellular localization is a function of nuclear export (Dong *et al.* 2009).

Therefore, in order to initially investigate whether or not *mier1*'s exon 3A-encoded 74bp insert does translate into a functional NES, MCF7 cells transfected with either pCS3+MT*mier1* $\alpha$ , pCS3+MT *mier1* $\alpha$  exon 3A, or pCS3+MT tagged control were treated with 10ng/ml of LMB for 8 hours, forty hours post transfection (refer to section 2.8.1 for methodology). Following the eight hour time point of LMB treatment, ICC was performed and cells were counted for staining. Like the previous section, the staining pattern (ie: cytoplasmic, nuclear, and whole cell staining) was determined.

Figure 3.17 shows the percentage staining for both isoforms following LMB treatment in MCF7 cells. MIER1 alpha was not affected by LMB treatment, as the percentage of stained cells did not significantly differ between the non-treatment and treatment groups ( $p>0.05$ ). MIER1 3A alpha staining in MCF7 cells did, however, change in response to the LMB treatment as 18% less cells displayed exclusively cytoplasmic staining following treatment when compared to the non-LMB treated control cells (Fig 3.17). This decrease in cytoplasmic staining of MIER1 3A alpha was followed by an increase in whole cell staining, whereby 8% more cells exhibited whole cell MIER1 3A alpha staining following LMB treatment than in the absence of LMB. Furthermore, nuclear staining also increased following LMB treatment, with 10% more cells displaying nuclear MIER1 3A alpha staining when compared to cells that were not treated with LMB (Fig 3.17).



**Figure 3.17: Effect of Leptomycin B treatment on *miel1* isoform subcellular localization in MCF7 cells**

MCF7 cells were seeded at a density of 15,000 cells/ well, in an 8-well chamber slide and grown in cell line specific supplemented DMEM for approximately 18 hours. Chamber slide wells were then transfected with 0.26µg of either pCS3+MT*miel1* $\alpha$ , pCS3+MT *miel1* $\alpha$  exon 3A, the MYC-tagged control vector (pCS3+MT), or were left untransfected. Forty hours following transfection 10ng/ml of LMB was added to each well (equal volume of 0.1% ethanol, which is the vehicle for the LMB, was added to non- treatment wells). Eight hours following LMB treatment (forty eight hours following transfection), cells were fixed with 4% paraformaldehyde and ICC was performed with an antibody specific for the MYC tag. Stained cells were counted and categorized as either having nuclear, cytoplasmic, or whole cell staining. The y axis represents the proportion of cells in each category whereby the x axis denotes the specific transfected *miel1* isoforms and the tagged control whereby the isoform produced from transfection with pCS3+MT*miel1* $\alpha$  is termed MIER1 alpha and the isoform produced from produced from pCS3+MT *miel1* $\alpha$  exon 3A transfection is termed MIER1 3A alpha (n=7 for all conditions except MIER1 3A alpha -LMB whereby n=6).



### **3.5.3 Subcellular Localization of *mier1* Isoforms derived from Transcripts either containing or lacking the 74bp Insert encoded by Exon 3A following Leptomycin B Treatment in HEK 293 cells and MDA MB 231 cells**

The previous subsection of this study analyzed the subcellular localization of MIER1 alpha and MIER1 3A alpha in conjunction with LMB treatment in a well established ER+ breast carcinoma cell line. In order to further investigate MIER1 $\alpha$  isoform localization in cell lines lacking ER signalling, the exact same experimental methodology (refer to sections 2.8 & 2.8.1) was utilized for HEK 293 cells and the ER- breast carcinoma cell line MDA MB 231. Figure 3.18 illustrates the proportion of transfected cells that displayed nuclear, cytoplasmic, or whole cell staining following -LMB and +LMB treatment in both HEK 293 cells (Fig 3.18 A) and MDA MB 231 (Fig 3.18 B).

The immunocytochemical staining, and thus subcellular localization, of both the MIER1 alpha isoforms in the absence of LMB treatment was prominently whole cell staining<sup>22</sup> for both cell lines independent of isoform type (90% whole cell staining for MIER1 alpha in HEK 293 cells, 89% whole cell staining for MIER1 3A alpha in HEK 293 cells, 90% whole cell staining for MIER1 alpha in MDA MB 231 cells, and 100% whole cell staining for MIER1 3A alpha in MDA MB 231 cells) (Fig 3.18 A & B). In HEK 293 cells, the amount of cytoplasmic staining for both isoforms was not significantly higher than the nuclear staining (7% of MIER1 alpha-transfected HEK 293 cells and 10% of MIER1 3A alpha-transfected HEK 293 cells portrayed cytoplasmic staining while 1% of MIER1 alpha- transfected HEK 293 cells and 3% of

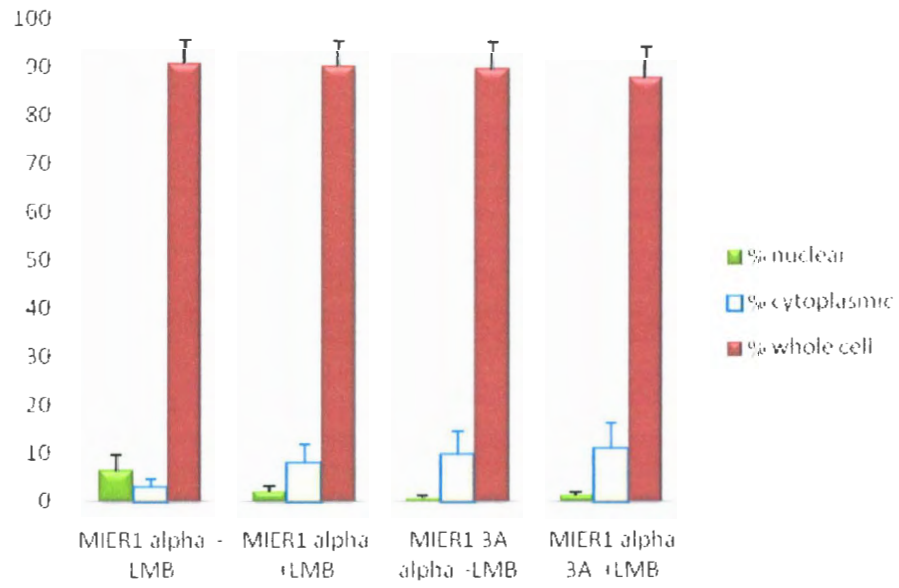
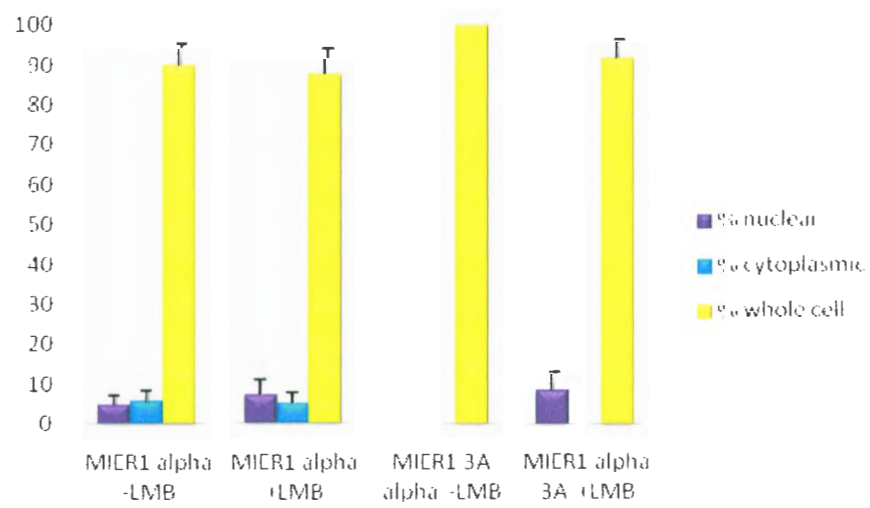
---

<sup>22</sup> p<0.001 when comparing whole cell staining to cytoplasmic and nuclear staining.

MIER1 3A alpha- transfected HEK 293 cells portrayed nuclear staining)<sup>23</sup> (Fig 3.18 A & B). Cytoplasmic and nuclear staining for MDA MB 231 cells was non-existent in MIER1 3A alpha-transfected MDA MB 231 cells and did not significantly differ among the MIER1 alpha-transfected MDA MB 231 cells with 4.7% of cells displaying cytoplasmic staining and 5.3% of cells displaying nuclear staining ( $p>0.05$ ) (Fig 3.18 A & B). It is also important to note that these aforementioned staining patterns and trends for both HEK 293 cells and MDA MB 231 cells were not significantly affected by LMB treatment ( $p>0.05$ ) (Figure 3.18 A & B).

---

<sup>23</sup>  $p>0.05$  when comparing nuclear staining to cytoplasmic staining among both MIER1 $\alpha$  isoforms.

**A****HEK 293 Cells****B****MDA MB Cells**

**Figure 3.18: Effect of Leptomycin B treatment on *mier1* isoform subcellular localization in HEK 293 cells and MDA MB 231 cells**

HEK 293 cells were seeded at a density of 12,000 cells/ well and MDA MB 231 cells at a density of 15,000 cells/ well in an 8-well chamber slide and grown in cell line specific supplemented DMEM for approximately 18 hours. Chamber slide wells were then transfected with 0.26 $\mu$ g of either pCS3+MT*mier1* $\alpha$ , pCS3+MT *mier1* $\alpha$  exon 3A, the MYC-tagged control vector (pCS3+MT), or were left untransfected. Forty hours following transfection 10ng/ml of LMB was added to each well (equal volume of 0.1% ethanol, which is the vehicle for the LMB, was added to non- treatment wells). Eight hours following LMB treatment (forty eight hours following transfection), cells were fixed with 4% paraformaldehyde and ICC was performed with an antibody specific for the MYC tag. Stained cells were counted and categorized as either having nuclear, cytoplasmic, or whole cell staining. The y axis represents the proportion of cells in each category whereby the x axis denotes the specific transfected *mier1* isoforms and the tagged control whereby the isoform produced from transfection with pCS3+MT*mier1* $\alpha$  is termed MIER1  $\alpha$  and the isoform produced from produced from pCS3+MT *mier1* $\alpha$  exon 3A transfection is termed MIER1 3A  $\alpha$ . Figure 3.18 A depicts results from HEK 293 cells (n=4), and Figure 3.18 B represents results from MDA MB 231 cells (n=4).



## Chapter 4: Discussion

Early manifestations of specific events leading to neoplastic development prime a cell to be resilient and noncompliant to the cellular restraints and security networks that are pivotal to ensuring functional cellular homeostasis. Such security restraints and networks are essentially regulatory pathways governing fundamental physiological processes that, when altered or dysregulated, promote survival, proliferation, and cellular growth (Fig 1.1, 1.2; Hanahan & Weinberg, 2000; Luo *et al.*, 2009). At the basic molecular level though, these alterations in regulatory signalling pathways that attribute a cancerous phenotype have a common converging point. This point of convergence is that of the fundamental process of gene expression whereby the production of every protein in the cell is tightly specified, controlled, and dictated by copious inter-related and inter-connected regulatory factors at multiple levels through a conglomerate of signalling pathways (sections 1. 2- 1.3).

Over the past decade, *mier1* has been well documented as playing a role in the regulation of gene expression, as it encodes a potent nuclear transcriptional regulator protein (Ding *et al.*, 2003; Ding *et al.*, 2004; Paterno *et al.*, 1997). *mier1* transcript variants exert a transcriptional repression function in the cell using various mechanisms, with one isoform in particular (MIER1 $\alpha$ ) acting as a corepressor of ER $\alpha$ . By virtue of MIER1's transcriptional regulatory role in the cell, it is no surprise that this protein has also been implicated in breast cancer, as the MIER1 $\alpha$  isoform was revealed to undergo an aberrant subcellular localization switch from the nucleus to the cytoplasm as breast carcinoma progresses to a more invasive state (McCarthy *et al.*,

2008). Furthermore, the overexpression of MIER1 $\alpha$  was shown to inhibit estrogen-induced anchorage independent growth in an ER+ breast carcinoma cell line, a result consistent with MIER1 $\alpha$ 's role as a corepressor of ER $\alpha$  (McCarthy *et al*, 2008).

In efforts to study the regulation of *mier1* in breast cancer, such as the specific regulatory elements controlling expression of this gene and the mechanisms affecting subcellular localization of its protein isoforms, this study first took a step back to investigate the regulation of *mier1* through a promoter activity analysis approach. As described in section 1.5.2, *mier1* has two distinct promoter regions, one of which, MAEP-P2, had already been characterized with regards to the precise nucleotide sequences involved in transcription initiation, as well as with regards to putative *cis*-acting elements that could act to recruit various transcription factors involved in the regulation of *mier1* expression (section 1.5.2; Ding *et al.*, 2004). *mier1*'s most 5' promoter, MLP-P1; however, had yet to be analysed prior to this study.

As outlined in this study's objectives (section 1.6), the principle goal of my Masters research project was to **characterize the MLP-P1 promoter** and to find the sequences involved in maximal and minimal activation at this promoter. Once this information had been established, it was then an additional aim to analyze and investigate various aspects of *mier1* promoter activity in breast cancer cells, such as: 1) whether the nucleotide sequences encompassing the MLP-P1 promoter region have any putative TFBSs that interact with TFs that are dysregulated in breast cancer; 2) whether or not MLP-P1 is increasingly or preferentially activated in breast cancer cells; 3) whether the ER status of the particular breast cancer cell could potentially influence MLP-P1 activity; 4) whether differential regulation persists at either of *mier1*'s two promoters in breast cancer cells; and 5) whether there are true



implications for preferential activity at MLP-P1, such as: does the putative NES and/or transmembrane domain act to remove MIER1 $\alpha$  from the nucleus? As alluded to previously though, the key prerequisite in order to investigate these latter additional issues surrounding *mier1* transcriptional regulation at MLP-P1 in breast cancer cells is the characterization of the sequences that are actively involved in eliciting transcription at *mier1*'s MLP-P1 promoter.

#### **4.1 *mier1*'s MLP-P1 Promoter Region and Promoter Activity in Breast Cancer Cells**

##### **4.1.1 The Minimal Promoter of MLP-P1**

*The minimal promoter of MLP-P1 lies between -91bp and -44bp from the MLP-P1 putative TSS in MCF7 breast carcinoma cells*

Results from the transient transfection of luciferase reporter gene deletion constructs harbouring putative promoter regions of MLP-P1 into MCF7 breast cancer cells revealed that, in these cells, the minimal promoter of MLP-P1 resides within -91bp of the putative TSS (Fig 3.2). The smallest MLP-P1 luciferase deletion constructs, MLP-P1 (-44) pGL3 and MLP-P1 (+37) pGL3, did not produce significantly higher levels of RLU above the pGL3 empty vector negative control ( $p>0.05$ ) (Fig 3.2). We expected that this would not elicit significant levels of promoter activity as the 5' starting nucleotides of this construct's insertion sequence reside 37bp downstream of the putative TSS of MLP-P1 and are thus situated in the first exon (Exon 1A) of *mier1*. Conventionally, active promoters are not intrinsic to exonic or intronic regions, and as such, these regions do not normally serve to recruit

transcription initiation machinery (Bajic *et al.*, 2006; Carninci *et al.*, 2006; Trinklein *et al.*, 2002). However, with that being said, some studies have found that promoter regions can exist in exons and/ or introns (Carninci *et al.*, 2006; Gaildrat *et al.*, 2005; Kaku & Rothstein, 2009; Singaraja *et al.*, 2001). Consequently, as both intronic and exonic promoter regions have been described, this study employed the pGL3 empty vector lacking any putative MLP-P1 promoter sequence upstream of the *luc* + gene as a negative control for promoter activity.

Interestingly, the construct MLP-P1 (-44) pGL3, which contains 44bp of sequence upstream of the putative MLP-P1 TSS (47bp less than the MLP-P1 (-91) pGL3 construct), also did not produce significantly higher levels of luciferase activity than the pGL3 empty vector. The significance of this finding is that it indicated that in MCF7 breast cancer cells, the nucleotide sequence between positions -91 and -44 contained the necessary *cis*-elements required to initiate transcription at this promoter, thereby enabling the generation of the reported luciferase activity levels. The nucleotide sequence downstream of the -44 position, and therefore the sequences cloned into MLP-P1 (-44) pGL3 that produced luciferase activity at the same level as the empty control vector, must not have contained ample *cis*-acting elements required to sufficiently recruit the essential transcriptional machinery needed to produce high levels of detectable luciferase activity. Refer to Figure 2.2 for the actual nucleotide sequence of this 47bp region (between the positions of -91 and -44) as well as the location of this particular sequence with regards to the MLP-P1 putative TSS. Moreover, the benefit of narrowing down this aforementioned sequence involved in minimal MLP-P1 promoter activity is that it can be furthermore analyzed for putative TFBSs, some of which may recruit TFs that could possibly be dysregulated



throughout cancer development & progression (refer to sections 3.3 & 4.3 for further discussion and analysis of the TFBSs found in this particular area of the MLP-P1 promoter region).

#### **4.1.2 Maximum Activity Promoter Region of MLP-P1**

*The maximum promoter of MLP-P1 lies between -185bp and -91bp from the MLP-P1 putative TSS in MCF7 breast carcinoma cells*

As previously described in sections 1.2.1.1 & 1.3.1.1, promoter activity levels are a function of the interplay between the inherent *cis*-elements within the nucleotide sequences of a given promoter region and the *trans*-regulatory factors (such as TFs and coregulators) that perpetuate signal transduction pathways in a given cellular environment at specific time points (Cao *et al.*, 2008; Trinklein *et al.*, 2003). Therefore, the difference between maximal promoter activity regions and minimal promoter activity regions of a specific promoter is dictated by the extent of this interplay and the *cis*-elements that constitute the promoter nucleotide sequences. (Kininis & Kraus, 2008; Levine & Tijan, 2003; O Barrera & Ren, 2006). Upon comparison of the luciferase activity levels produced following transient transfection of MLP-P1 luciferase reporter gene deletion constructs, Figure 3.2 illustrates that the maximum activity promoter region of MLP-P1 resides between -185 and -91 bp from the putative MLP-P1 TSS in MCF7 breast cancer cells. As this data demarks this 94bp region between the -185 and -91 positions as constituting sequences resulting in the maximum activity at MLP-P1, it also further suggests that encompassed in this particular 94bp sequence are certain *cis*-elements that are more actively and

stringently involved in the recruitment of the transcriptional machinery than the apparent elements residing in the minimal MLP-P1 promoter region (between nucleotides -91 and -44 with reference to the MLP-P1 TSS). Refer to Figure 2.2 in order view the exact 94bp sequence between the MLP-P1 (-185) pGL3 and the MLP-P1 (-91) pGL3 construct 5' sequence insert starting positions.

Furthermore, the investigation for further distal-acting enhancer or repressor elements in the MLP-P1 promoter upstream region showed that a repressor element may exist between -1708 and -1077 bp upstream of the MLP-P1 putative TSS. Figure 3.2 demonstrates that MLP-P1 (-1708) pGL3, the largest engineered MLP-P1 luciferase reporter gene deletion construct, had significantly lower luciferase activity levels than MLP-P1 (-185) pGL3 ( $p < 0.001$ ) and comparable RLUs to MLP-P1 (-91) pGL3 ( $p > 0.05$ ). This data suggests that putative repressor elements, rather than further activation or enhancer elements, are present in these upstream, distal sequences. Although this particular study only continued to investigate and focus on the proximal promoter region of MLP-P1, the potential presence of a repressor element implicates the necessity of further study and investigation of these regions.

Primary future studies investigating this distal downstream repressive phenomenon could employ further MLP-P1 luciferase reporter gene deletion constructs that divide the 631bp region between MLP-P1 (-1708) pGL3 and MLP-P1 (-1077) pGL3 (a construct that generated RLU levels comparable to MLP-P1 (-185) pGL3) into additional segments. The sequences found to be highly involved in the generation of the reported decrease in luciferase activity could be further analyzed for putative *cis*-elements and repressor sequences via *in silico* examination. Subsequent



steps that could be utilized in order to investigate whether or not any of the putative TFBSs found are implicated in the regulation of this distal upstream region are discussed at the end of section 4.4.

#### **4.2 Effect of ER Status on MLP-P1 Promoter Activity in Breast Cancer Cells**

As briefly introduced in section 1.1.1, the complexity of cancer is not easily rivalled, as this disease is considered one of the most complex and intricate phenomena in biology (Edelman *et al.*, 2008; Grizzi & Chiriva-Internati, 2006). Even among the vast plethora of individuals that develop a given type of cancer, there exists multitudes of different subtypes; each subtype of which can even have unique qualities and attributes that manifest differently from individual to individual (Grizzi *et al.*, 2006). In breast cancer specifically, an important factor that is used to distinguish between the predominant mammary neoplasia subtypes is that of ER status (Lopez-Tarruella & Schiff, 2007). Accordingly, the importance of ER's state in developing breast tumours stems from its powerful predictive ability with regards to the success of breast cancer's mainstay therapy (endocrine therapy) (Clark, 1995; Grizzi *et al.*, 2006; Gruvberger *et al.*, 2001; Osborne, 1998). Recently, the heterogeneous nature of the ER subtypes has become even more and more evident, as tumours have been shown to present widely different percentages of ER+ cells that express the receptor protein at mixed intensities (Lopez-Tarruella & Schiff, 2007). Furthermore, ER+ and ER- tumours display remarkably different gene expression phenotypes that are not solely explained by differences in estrogen responsiveness, thereby highlighting the dynamic differences in the cellular environments developed

in these neoplastic subtypes (Gruvberger *et al.*, 2001; Lopez-Tarruella & Schiff, 2007).

In order to enhance the understanding of MLP-P1-specific promoter activity in breast cancer cells, this study addressed the issue of varying breast cancer cell ER status by investigating whether MLP-P1 promoter activity differs significantly in breast cancer cells either containing or lacking ER. Intriguingly, cell lines with the same ER status (ER- cell lines MDA MB 231 and Hs578T) generated significantly different levels of luciferase activity ( $p < 0.001$ ) from the maximum promoter activity producing construct MLP-P1 (-185) pGL3 (Fig 3.5). Specifically, MDA MB 231 cells produced 33 times less luciferase activity from MLP-P1 (-185) pGL3 than did Hs578T cells (Fig 3.5). This suggests that there must be other elements inherent to these particular cell lines that dictate the extent of the overall MLP-P1 promoter activity.

Conversely, the MLP-P1(-185) pGL3 luciferase reporter gene construct transfected into ER+ cell lines (MCF7 and T47D cells) generated comparable levels of RLU (Fig 3.5). This suggests that the regulating factors influencing the overall promoter activity level of MLP-P1 must either be present in both these cell lines, or cell line-specific regulatory factors must induce transcriptional activation at MLP-P1 to a similar extent. Moreover, Figure 3.5 also demonstrates that RLU levels generated from the MLP-P1 (-185) pGL3 construct transfected into the ER- cell line Hs578T produced analogous levels of luciferase activity to the ER+ cell lines. This furthermore suggests that this ER- cell line may contain common regulatory elements and activated signalling pathways as the employed ER+ cell lines.



The key limitation with regards to this latter type of comparison across different cell lines is that even though the ER status of each cell line is known, the status of the other MLP-P1 putative regulatory pathways and elements in these cell lines is not entirely characterized. This is mainly because *mier1* specific transcriptional regulation has not been fully analyzed to date, and also because different cell lines constitute distinct transcriptomes and have diverse cellular environments encompassing varying activated signalling pathways. As noted in Table 2.1, the exact origins and even level of cancer progression for each of these cell lines vary. This adds to the possible variation of each respective cellular environment composition and the exact activated signalling networks in these cells and; therefore, cell lines lacking ER differ from ER+ cells in more respects than merely the lack of this receptor (Angus *et al.*, 1999). For example, T47D cell lines portray classic epithelial cell-like morphologies, while the MDA MB 321 cells are typically fibroblast-like and have been demonstrated to behave in a more advanced malignant and metastatic nature (Angus *et al.* 1999). These cell lineages thus introduce differences in many additional regulatory processes, such as growth factor and plasma membrane receptor expression (refer to Angus *et al.*, 1999 for a further detailed review on further differing attributes among the commonly established human breast cancer cell lines).

Furthermore, these different cell lines also each have distinct transfection efficiencies that dictate the extent of MLP-P1 luciferase reporter gene deletion construct uptake. This study controlled for transfection efficiency by co-transfecting a beta galactosidase expressing vector with the luciferase reporter gene deletion constructs. An obstacle with this particular control, however, is that one cannot be

certain that the luciferase reporter gene deletion constructs are being transfected exactly the same extent as the beta galactosidase vector or whether they are both even entering the same cell. In order to verify whether this is truly a confounding variable, these experiments could be repeated using a dual-luciferase reporter system (Promega: Dual-Luciferase® Reporter Assay System: Technical Manual – Instructions for use of products E1910 and E1960). This particular reporter system is designed to contain two reporter enzymes in a single system with each reporter vector constituting an experimental reporter enzyme (firefly luciferase, which would indicate the extent of promoter activity) and a control reporter enzyme (renilla luciferase, which would serve as the baseline response) (Promega). Measurements from the experimental reporter and the internal control reporter would then not only be originating from the same cell, but would stem from the exact same reporter vector.

Consequently though, as it is clearly evident that there are numerous factors that contribute to the heterogeneity among these breast cancer cell lines, it cannot directly be deduced through these studies that the differences in MLP-P1 promoter activity obtained are absolutely correlated with the cell line-specific ER status alone. The reason these specific experiments were employed though, was so that reporter activity data could be compared across breast cancer cell lines differing in ER status and serve as preliminary support for the possible implications of ER status on the activation at the MLP-P1 promoter in breast cancer cells. In order to firmly and accurately investigate whether or not estrogen, and consequently ER signalling, affects activation at this particular promoter, experiments comparing the effect of estrogen within the same cell line either containing or lacking ER would be more

conducive to narrowing down the exact functional implications of ER signalling on the activity at a particular promoter.

This type of comparison experiment can be performed using various approaches. One possible approach could be by knocking down ER expression using short interfering RNA (si -RNA). These si-RNA-ER specific treated cells could then be compared to the same cell line that was treated with a control si-RNA. Some groups have even been known to use ER inhibitors (such as the anti-estrogen ICI 182, 780) in order to analyze the affect of ER signalling in various ER+ cell lines (Dauvois *et al.*, 1993; Fawell *et al.*, 1990). Other research groups, in order to minimize differences in cell lineage, have even developed stable clones of already established cell lines with a new ER status instead of transiently changing the ER status. For example, in efforts to analyze the effect of ER on activity of cytochrome P4501B1 (CYP1B1) and cytochrome P4501A1 (CYP1A1) across a variety of breast cancer cell lines, Angus *et al.*, 1999 developed paired cell lines deriving from the same background, but which differed in ER status (Angus *et al.*, 1999). Furthermore, an ER- clone T47D:C4:2W (T47D-) had previously been derived from normal ER+ T47D cells by long-term culture in estrogen-free medium (Pink *et al.*, 1996). These cell lines could then be used for further promoter activity analysis as performed in this current study.



***MLP-P1 promoter activity patterns of relative luciferase activity levels remained constant when compared between MLP-P1 luciferase reporter gene deletion constructs in all examined cell lines independent of ER status***

Finally, even though this set of data was unable to fully reveal whether or not ER status truly affects transcriptional activation at MLP-P1, there was one very important finding that these experiments did exhibit. This is that across all the examined breast cancer cell lines, the relative promoter activity (measured as relative luciferase activity levels) stemming from each MLP-P1 luciferase reporter gene deletion construct remained constant when the activities of these constructs were compared within a particular cell line (Fig 3.4). This invariable pattern, which was observed independent of ER status, was first characterized in section 3.1 whereby the MLP-P1 (-185) pGL3 construct consistently generated the highest level of promoter activity in each cell line with the other MLP-P1 luciferase reporter gene deletion constructs consecutively decreasing in RLU levels as the size of the putative MLP-P1 promoter insert shortened (Fig 3.2 & 3.4). This phenomenon insinuates that these areas of the MLP-P1 promoter are activated in a similar manner with respect to each other across breast cancer cell lines, and therefore demonstrates that there is a commonality with regards to the response of the sequence areas of the MLP-P1 proximal promoter in breast cancer cells. In conclusion, with regards to MLP-P1 promoter activity in breast cancer cells, that which varies between breast cancer cell lines with respect to MLP-P1 promoter activity is the intensity or efficiency by which these areas of the MLP-P1 promoter can recruit the transcription initiation machinery. Relatively though, within a given cell line, the specific sequences of the MLP-P1 promoter are similarly activated with respect to one another, thereby generating consistent promoter activity patterns (Fig 3.4).



### **4.3 MLP-P1 Promoter Activity in Breast Cancer Cells vs. the Non-Cancerous HEK 293 Cell Line**

Over the past few decades, breast cancer research has revealed numerous genes that are either associated or dysregulated throughout mammary neoplastic development (Polyak, 2007). In order to initially investigate as to whether *mier1*'s MLP-P1 promoter is differentially regulated in a breast cancer cellular environment compared to a non-cancerous cellular environment, data from reporter gene analysis experiments following transfection of MLP-P1 luciferase reporter gene deletion constructs was compared between the non-cancerous cell line HEK 293s and an assortment of breast cancer cell lines (MDA MB 231, Hs578T, MCF7, and T47D cells). The HEK 293 cell line generated significantly lower levels of luciferase activity than the ER+ breast cancer cell lines and the ER- Hs578T breast cancer cell line (Fig 3.6). This observation supplies preliminary support that MLP-P1 is activated at a higher level in these breast cancer cells than in non-neoplastic human embryonic kidney cells. Even more interestingly though, was that HEK 293 cells had higher promoter activity levels than the ER- MDA MB 231 cell line. This suggests that the extent of activity at this promoter in MDA MB 231 cells is even lower than that of the non-cancerous HEK 293 cells, thereby further indicating that the signalling environment developed in this cell line is not conducive to increasing MLP-P1-mediated promoter activity.

This type of comparison though, because it similarly uses analytical methods discussed in section 4.2, also succumbs to the same pitfalls. Such pitfalls are that we can only primarily hypothesize and not directly deduce that the lack of a neoplastic milieu, as is the case with the HEK 293 cells, is the only reason for these differences in overall promoter activity levels at the MLP-P1 promoter. Other factors that are specifically inherent to embryonic kidney cells, but not to mammary epithelial cells, may potentially function together in order to give the observed results. Ideally, in order to accurately study and investigate whether or not MLP-P1 is being differentially activated in cancer vs. normal cells, one would use the same type of cells (ie: mammary epithelial cells); one population which would constitute normal mammary epithelial cells, and another population composed of cancerous mammary epithelial cells of the neoplastic phenotype.

To date, various “normal-like” mammary epithelial cell lines have been developed, such as: MCF 10, MCF 10A, MCF 10F, MCF 12, etc... ([www.atcc.org](http://www.atcc.org)). The term ‘normal-like’ is required for the description of these cell lines as they cannot truly be classified as ‘normal’ cells due to either the original source of the cells or cellular adaptations following immortalization- a necessary step in the establishment of cell lines (Singhal *et al.*, 1999). For example, MCF 10 cells do not have a normal karyotype consequent to immortalization and, furthermore, were collected from a patient presenting with fibrocystic disease following a mastectomy (Singhal *et al.*, 1999). However, in spite of these characteristics, this particular cell line is considered a normal mammary control and a good model to study the carcinogenic process using chemical carcinogens and oncogenes for induction of malignant transformation (Hu *et al.*, 1998; Ismail *et al.*, 1999; Wang *et al.*, 2000; Wei *et al.* 1998).

Additionally, the Hs578T ER- breast cancer cell line could also be used for the purpose of evaluating MLP-P1 promoter activity between cancerous and non-cancerous mammary epithelial cells by comparing Hs578T-specific experimental results to its non-tumourigenic counterpart cell line, Hs578BsT. Hs578BsT was derived by Hackett *et al.* 1977 from normal breast tissue peripheral to the infiltrating ductal carcinoma source for Hs 578T (Hackett *et al.* 1977). Consequently, future studies could compare results following MLP-P1 reporter gene deletion construct transfection from the non-tumourigenic 'normal-like' mammary epithelial cell lines to developed breast carcinoma cell lines. Using this same cell type and/ or cell origin diminishes cell-type specific variability in result interpretation.

***HEK 293 cells exhibited similar promoter activity patterns when compared between the MLP-P1 luciferase reporter gene deletion constructs, just as the previously analyzed breast cancer cells.***

Finally, even though HEK 293 cells resulted in different overall MLP-P1 luciferase activity levels when compared to multiple breast cancer cell lines, the HEK 293 results did recapitulate the same MLP-P1 promoter activity patterns as the breast cancer cells (Fig 3.5). This pattern, as first described in section 3.1, illustrates that the nucleotide sequence between -185 and -91 elicits the highest amount of activity followed sequentially by the other MLP-P1 luciferase reporter gene deletion constructs in order of insert size. The significance of this observation in these particular cells is that it shows that this activity pattern is present regardless of cell type and lineage, thereby further highlighting the functional importance of these



nucleotides and their putative regulatory role in the generation of MLP-P1 specific *mier1* transcripts.

A final point to note is that the HEK 293 cells used for these experiments were transfected with ER $\alpha$  along with the MLP-P1 and/ or MAEP-P2 luciferase reporter gene deletion constructs and the  $\beta$ -gal expression vector. HEK 293 cells intrinsically do not express ER $\alpha$ ; therefore, this latter step served as a control to ensure that the obtained results were not affected by a lack in expression of this nuclear hormone receptor. Due to this particular control, ER+ breast cancer cell lines and HEK 293 cells results can be more directly compared as they have one less confounding variable. However, the principal overall confounding variable is still present as HEK 293 cells and MCF7 cells are not only completely distinct cell lines, but are as well completely distinct cell types.

The capacity to control the ER status in HEK 293 cells through transient transfection though provides a great opportunity to further investigate whether or not ER status can affect the overall level of promoter activity at MLP-P1, as one can observe MLP-P1 activity in the presence and absence of ER in the same cell line. Therefore, future experiments with these cells could repeat this same type of experimental procedure, but also have a set of cells lacking transient transfection with ER $\alpha$ . In this case, as results containing or lacking ER would stem from the same cell line, the investigator would be able to efficiently investigate the effect of ER on MLP-P1 promoter activity in these cells.



#### **4.4 Putative Transcription Factors involved in the Regulation of *mier1* at the MLP-P1 Promoter**

What is truly causing these aforementioned alternating overall MLP-P1 promoter activity levels across different breast cancer cell lines and the non-cancerous HEK 293 cell line, though? As described in sections 1.2 & 1.3, the expression of a particular gene is dictated by the interaction between the inherent *cis*-elements of the gene's promoter region and the active transcriptional regulators present in the cellular environment (Frith *et al.*, 2008; O Barrera & Ren, 2006). As the cellular environments, and therefore the level of various activated TFs, coregulators, and signalling pathways of each respective cell line differ, the interplay between these regulatory molecules and the *cis*-acting response elements residing in a given gene promoter region changes in correlation. Therefore, one would expect the ensuing differences in cell line-specific environments to mediate distinct signalling networks, some of which may diversely affect gene transcriptional activation via differential interaction with the gene promoter, and especially with regard to this project, the MLP-P1 promoter.

***Transcription factors such as Sp1, CREB, C/EBP $\beta$ , USF, and c-Rel may play a pivotal role in MLP-P1 specific transcriptional regulation.***

In order to fully understand how the MLP-P1 promoter is being affected and/or regulated in breast cancer cells, it is not only important to investigate the various signalling pathways and TFs that are activated, but as well to delineate the *cis*-response elements residing in the particular sequences involved in generating

promoter activity at this promoter. This study demonstrated that the nucleotide sequence spanning position -185 to -44 from the MLP-P1 putative TSS are involved in promoting transcriptional activation of *mier1* from MLP-P1 in breast cancer cells (Figures 3.2 & 3.4, sections 3.1 & 3.2). Encompassed in this sequence was an area involved in eliciting maximal promoter activity (nucleotides spanning -185 to -91) and a region harbouring the nucleotides necessary for minimal promoter activity (nucleotides spanning -91 to -44) (Fig 3.2). Table 3.1 demonstrates that Sp1 and CREB are the main candidate transcription factors affecting the maximal promoter of MLP-P1 and that Sp1, C/EBP $\beta$ , USF, and c-REL are candidate transcription factors activating the minimal MLP-P1 promoter region. The following subsections outline these TFBSs and discuss whether or not the particular TF that binds to the established putative MLP-P1 specific TFBS is dysregulated in cancer and/ or has a role in breast cancer progression specifically.

### **1) Specificity Protein 1 (Sp1)**

Putative TFBSs for this particular transcription factor were found in both the maximal and minimal promoter activity producing areas of MLP-P1 (Fig 3.8). Interestingly this transcription factor, which binds to the GC rich core consensus sequence of 5'-GGCGGG-3 (Risili *et al.*, 1995), was also characterized as regulating transcription from *mier1*'s MAEP-P2 promoter, a function that was even shown to be negatively regulated by MIER1 in a negative feedback loop mechanism (sections 1.5.2.1 & 1.5.3.2.1; Ding *et al.*, 2004). The presence of Sp1 sites in the MLP-P1 active promoter region further insinuates that MIER1 could further negatively repress transcription at MLP-P1, as well as MAEP-P2.

In terms of the neoplastic cellular environment, Sp1 has been intimately linked to growth-related changes in transcription (Kim *et al.*, 2005). There is also emerging evidence that Sp1 protein expression may be a critical factor in tumour development, growth, and metastasis (Safe & Abdelrahim, 2005). Moreover, Sp1 has been found to maintain cell proliferation through various mechanisms. One such mechanism is by sustaining expression of fatty acid synthase (FAS), which allows cancer cells to constantly biosynthesize their plasma membrane, a step critical to continual cell propagation (Lu & Archer 2009). Another mechanism is that Sp1 also plays a role in regulating cell cycle via regulation of CDC25A expression, which is a pivotal cell cycle checkpoint mediator and regulator of both the G1/S and G2/M transitions (Ray *et al.*, 2007).

In the case of breast cancer specifically, one study showed that Sp1 protein expression was elevated in 11 out of 14 breast carcinomas, whereas only 1 in 5 benign breast lesions expressed detectable Sp1 protein levels (Black *et al.*, 2001). Furthermore, an additional recent study demonstrated, by analyzing normal breast and breast cancer tissues and specimens, that Sp1 staining in cancer tissue was positively correlated to TNM stage, tumour invasion, and lymph node metastasis. This group specifically proposed that Sp1 maybe participate in the invasion and metastasis of breast cancer, and is one of the valuable markers indicating poor prognosis of breast cancer (Wang *et al.*, 2007).

Finally, another principle interest with regards to Sp1's role in breast cancer progression is its interactive synergy with the ER $\alpha$  receptor, whereby ER $\alpha$  has been found to tether to the Sp1 TF and activate transcription following the binding of the



ER $\alpha$ /Sp1 complex to cognate Sp1 sites. This pivotal interaction can even occur in the absence of ligand (Kim *et al.*, 2005; Porter *et al.*, 1997).

## **2) cyclic AMP (cAMP) Response Element Binding Protein (CREB)**

A CREB binding site [cyclic AMP (cAMP)-response element (CRE)] was the second key putative TFBS that was found in between positions -185 to -91, and thus may potentially be involved in eliciting maximal activity at MLP-P1 (Fig 3.7). CREB is a member of a family of transcription factor proteins involved in stimulus-transcription coupling, and as such generally functions as a classic second messenger (Mayr & Montminy, 2001). More specifically, CREB-mediated transcription regulates genes and processes involved in glucose homeostasis, growth-factor-dependent cell survival and growth (Mayr & Montminy, 2001). Interestingly though, this TF is best characterized in the brain where it is best known for its roles in learning and memory to which it mediates experience-based neuroadaptations (refer to the comprehensive review by Carlezon Jr *et al.*, 2005).

Various roles for CREB activity in cancer development and signalling pathways though have also been revealed following investigation into the regulatory mechanisms behind aromatase expression, as well as pathways preventing cellular migration (Ghosh *et al.*, 2008; Hansen *et al.*, 2009). In the case of aromatase, which is the rate-limiting enzyme in estrogen biosynthesis and a key target in breast cancer treatment, Ghosh *et al.* 2008 demonstrated that CREB indirectly regulates its expression (Ghosh *et al.*, 2008). They revealed that CREB actually negatively regulates aromatase basal expression by maintaining constitutive BRCA1 expression



levels and proposed that inhibition of CREB activity would decrease BRCA1 expression thereby leading to elevated aromatase basal expression. In terms of cellular migration, Hansen *et al.*, 2009 showed that Wnt-5a induces a cAMP response leading to Thr34 phosphorylation of DARPP-32 and a subsequent downstream activation of CREB. This CREB-specific activation resulted in the inhibition of breast cancer cell migration (Hansen *et al.*, 2009).

Mechanistically, CREB mediates transcriptional regulation by dimerizing and binding to the cyclic AMP (cAMP)-response element (consensus sequence: TGACGTCA) and thereby inducing transcription of its downstream targets (Carlezon *et al.*, 2005; Mayr & Montminy, 2001). Molecular studies of this activation pathway have revealed that phosphorylation, and subsequent activation of the CREB TF, recruits a cascade of associated proteins such as CREB-binding protein (CBP), which in turn promotes the assembly of a larger transcriptional complex involved in altering the conformation of nearby chromatin thereby enabling RNA synthesis via RNA polymerase II (Bannister *et al.*, 2002; Carlezon *et al.*, 2005; Walker *et al.*, 1996).

The presence of a putative CRE in the MLP-P1 promoter region is further interesting in the case of *mier1* regulation, as it was previously discovered that the MIER1 $\beta$  isoform inhibited CBP and its HAT activity (Blackmore *et al.*, 2008). This suggests that perhaps *mier1* MLP-P1 specific transcription could be involved in a type of negative feedback mechanism towards CREB-mediated transcriptional activation. Further studies with regards to such a role in regulating this specific pathway are needed.

### 3. The CCAAT/Enhancer Binding Protein Beta (C/EBP $\beta$ )

C/EBP $\beta$  is a member of a family of basic leucine zipper (bZIP) transcription factors which bind specific DNA sequences as either homo- or heterodimers (consensus C/EBP=  $_{A/G}TTGCG_{C/T}AA_{C/T}$ ) (Grimm *et al.*, 2003). This study's *in silico* analysis of the functional regions of MLP-P1 found that such a TFBS resides between -91 and -44bp upstream of the putative MLP-P1 TSS (Fig 3.7). The C/EBP family of TFs in particular have been implicated in regulating transcription and functioning of various critical cellular processes such as cellular proliferation, differentiation, survival and/or apoptosis, metabolism, inflammation, and oncogene-induced senescence and tumorigenesis (Zahnow, 2009). The study of C/EBP regulatory effects in the cell has been quite cumbersome though, as in general these transcription factors exert pleiotropic effects based on many factors such as: tissue- and stage-specific gene expression, alternative translation of various protein isoforms, interaction with other transcription factors, and variable DNA-binding specificities (Zahnow, 2009).

Furthermore, the C/EBP $\beta$  isoforms have been reported as playing distinct and important roles in the development of many tissue types, and as such, have also been associated with the neoplastic state (Grimm *et al.*, 2003). With respect to breast cancer in particular, C/EBP $\beta$  mRNA levels are mostly non-elevated compared with normal tissue, but are increased in a more-aggressive subset of tumours versus the less-aggressive tumours (Grimm *et al.*, 2003). Moreover, significant increases in the LIP: LAP ratio (-a ratio of two specific C/EBP $\beta$  isoforms) have been observed in estrogen receptor-negative, aneuploid, and highly proliferative breast tumours that are

associated with a poor prognosis (Zahnow, 2009). Additionally, some researchers attest that C/EBP $\beta$ 's role in cancer may be due to its mediated role in regulating cell survival and apoptosis. Studies to this effect, however, have showed conflicting results which further illustrate this TF's complex signalling characteristics, as it has the potential to promote the survival of some transformed cells while inducing growth arrest in others (Zahnow, 2009).

#### **4. Upstream Stimulatory Factor (USF)**

The TFBS for this TF was also found in the MLP-P1 proximal promoter region spanning -91bp to -44bp upstream of the MLP-P1 putative TSS (Fig 3.7). There are two distinct USF genes in humans that encode for the USF-1 and USF-2 TFs respectively, which are both members of the eukaryotic evolutionary conserved basic-Helix-Loop-Helix-Leucine Zipper transcription factor family (Corre & Galibert, 2005). Downstream targets of these transcriptional regulators have been implicated in various key cellular processes, such as: stress and immune responses, cell cycle and proliferation, and lipid and glucid metabolism (Corre & Galibert, 2005). Molecularly, USF TFs interact with DNA as dimers and recognize E-box elements characterized by a central CACGTG or CACATG sequence, an action of which has also been shown to mediate recruitment of chromatin remodelling enzymes to the transcription pre-initiation complex (Corre & Galibert, 2005; Corre & Galibert, 2007; Ismail *et al.*, 1999; North *et al.*, 1999). Furthermore, they display strong similarities with the Myc oncoproteins both in their overall protein structure and DNA-binding specificity (reviewed in Ismail *et al.*, 1999).



Interestingly though, in terms of possible implications with carcinogenesis, USF proteins have been found to exert antiproliferative effects whereby overexpression of USF was demonstrated to prevent Ras and c-MYC- dependent cellular transformation and the proliferation of various transformed cells (Corre & Galibert, 2007; Ismail *et al.*, 1999). Furthermore, loss of USF function is common to not only breast cancer cells, but as well as in prostate hyperplasia and carcinoma (Chen *et al.*, 2006; Corre & Galibert, 2007; Ismail *et al.*, 1999). Moreover, distinct tumour suppressor genes, including p53, BRCA2 and APC, have also been shown to be direct targets of the USF TFs (Corre & Galibert, 2005).

## **5. c-Rel**

c-Rel is one of five subunits of the TF NF- $\kappa$ B (Gaspar Pereira & Oakley, 2008). This study found that a putative binding site for this particular subunit resides between the MLP-P1 minimal promoter region -91bp and -44bp. The significance of the appearance of this particular TFBS in the MLP-P1 promoter stems from the fact that NF- $\kappa$ B is a critical regulator of many cardinal cellular processes, and more particularly has been implicated in and contributes to each of the hallmarks of cancer development described in section 1.1 (Naugler & Karin, 2008; Ravi & Bedi, 2004). Its regulatory influence spans over 200 genes involved in these pertinent pathways, as countless studies have and still are investigating its functional role and downstream targets in the context of neoplastic development (Gaspar Pereira & Oakley, 2008; Naugler & Karin, 2008). Furthermore, some research groups claim that NF- $\kappa$ B may be highly involved in conferring resistance of human cancers to apoptosis as it is



frequently activated consequent to various genetic aberrations and the subsequently developed stress stimuli (Ravi & Bedi, 2004).

c-Rel in particular is a proto-oncogene first identified as the cellular counterpart of the v-Rel avian reticuloendotheliosis retrovirus (REV-T) derived oncogene (Liou & Hsia, 2003). This subunit is furthermore one of three NF- $\kappa$ B subunits that contains a *trans*-activation domain thereby facilitating direct gene transcription activation (Gaspar Pereira & Oakley, 2008). With respect to a role for this specific subunit in tumourigenesis, many human breast cancer tissue samples and derived cell lines express abnormally high levels of c-Rel and activation of this subunit has even been shown to be involved with mesenchymal to epithelial transition induction in mammary cells (Belguise *et al.*, 2007). Moreover, along with the cooperation of CK2A, a ubiquitously expressed serine/threonine kinase, c-rel has been shown to promote a more invasive gene profile *in vitro* (Belguise *et al.*, 2007).

With the finding of these putative TFBSs in the sequences involved in activity at the MLP-P1 promoter, future studies should now look at confirming a role for the respective TFs in the regulation of *mier1* at MLP-P1. More specifically, putative TFBSs could be mutated using site-directed mutagenesis. These 'mutated' MLP-P1 promoter sequences could then be cloned and inserted into the pGL3 empty vector thereby creating new MLP-P1 luciferase reporter gene deletion constructs; the experimental data from which could be compared to results collected from the non-mutated MLP-P1 luciferase reporter gene deletion constructs. For further confirmation following these MLP-P1 luciferase reporter gene promoter analysis experiments, additional experiments employing either transient transfection of vectors expressing

one of the six candidate TFs, or treatment with compounds or molecules that would activate the TF-specific signalling pathway, could be performed in conjunction with assays from the MLP-P1-specific luciferase reporter gene system. Levels of MLP-P1 activity could then be compared across the different treatment groups, for example: cells transiently expressing the TF vs. cells not expressing the TF. Western blots, would also be necessary following the transient transfection experiments in order to verify that the vectors were adequately and efficiently expressing the TF of choice. Additionally, if the cell line being studied endogenously expressed the TF in question, one could even knock down the particular TF using si-RNA. In conjunction with the si-RNA induced knockdown of the TF expression, luciferase reporter gene assays could then detect MLP-P1 promoter activity in the absence or presence of the si-RNA mediated TF specific knockdown.

Another type of approach that could be used that does not employ reporter gene technology is that of electromobility shift assays (EMSAs). Such an approach would assist in determining whether or not the candidate TF binds to the MLP-P1 promoter region. This specific method was utilized previously when investigating how Sp1 affected transcription at MAEP-P2 (Ding *et al.*, 2004). Briefly, these particular EMSAs would use a P<sup>32</sup> labelled probe specific to either the maximal or minimal MLP-P1 proximal promoter region. This probe would then be incubated with nuclear cell extracts. These extracts would then be stained with an antibody specific for the TF and the event of a shift due to antibody binding would be recorded. Moreover, the CHIP assays would be another very efficient method for revealing whether or not the candidate TFs do actually bind to the putative TFBSs found at the MLP-P1 proximal promoter.



Finally, in addition to analyzing whether or not the candidate TFs affect MLP-P1 promoter activity, and whether or not they are capable of binding to the MLP-P1 promoter, one could also verify whether or not the TF affects endogenous *mier1* expression. In this case, the particular cell line in question could be transiently transfected with expression vectors specific for the candidate TF. Forty-eight hours following transfection, total RNA could be isolated. Reverse transcriptase (RT) PCR could then be performed followed by either PCR or quantitative PCR (qPCR) with primers specific for the various *mier1* transcripts. The significance of these experiments is that they would demonstrate whether or not endogenous *mier1* transcription is affected by the particular TF. Furthermore, as all 12 possible transcripts stemming from *mier1* are characterized (Paterno *et al.*, 2002), transcript-specific primers could be utilized in order to monitor the effect of the candidate TF on each respective transcript. This would furthermore distinguish at which promoter the TF in question is acting, thereby furthermore verifying the MLP-P1 putative TFBS function, as transcriptional activation at both MLP-P1 and MAEP-P2 produce very distinct *mier1* transcripts.

Information obtained from such experimental assays validating the functional effects of the candidate TFs on MLP-P1 promoter activity would be markedly useful in delineating the regulatory network of *mier1*. All of the candidate TFs have been found to be either involved in or dysregulated throughout cancer growth and progression (with the exception of USF, which demonstrates antiproliferative and tumour suppressor properties). This suggests that MLP-P1 has the potential to become activated aberrantly throughout cancer progression, as these TFs may be abnormally activated or regulated. Of additional interest is with regards to the TFBSs for Sp1 and

CREB found in the MLP-P1 maximal promoter area (between -91 and -44bp from the putative MLP-P1 TSS). Both of these transcription factors have not only been found to contain implicated roles in breast cancer, but as well have been previously found to either be involved in *mier1* regulation (-with Sp1 already established as regulating *mier1* transcription at MAEP-P2) or affected by MIER1 function (-with CREB requiring CBP for downstream signalling and transcriptional activation; CBP being a molecule harbouring HAT activity to which MIER1 $\beta$  inhibits) (Blackmore *et al.*, 2008). Therefore, if these TFs do regulate MLP-P1 specific transcription, then further regulatory negative feedback loops involving MIER1 isoforms could possibly be revealed.

#### **4.5 MLP-P1 Promoter Proximal Sequence Conservation Across Various Species**

*Sequence alignment comparison of the human MLP-P1 promoter region to various species further reinforces functional importance of the MLP-P1 promoter proximal region.*

Sequences that mediate gene expression tend to be evolutionarily conserved (Loots, 2008). Consequently, data from inter-species genomic sequence comparison alignments are primed to aide in the decoding and identification of the sequences responsible for gene regulation (Pennacchio & Rubin, 2001). Out of the six species analyzed, MLP-P1 proximal promoter regions were highly conserved among four: rhesus (chimpanzee), mouse, rat, and dog (section 3.3.2; Appendix 7). The fact that there is such a high degree of conservation amongst these species furthermore reinforces the functional importance of these nucleotides.



#### 4.6 MLP-P1 vs. MAEP-P2 Promoter Activity

The regulatory influence that alternate promoter usage exerts on gene expression is becoming more and more apparent, not only because of the increased discovery of genes that are regulated at more than one promoter, but due to the further characterized role of alternate promoter usage in tissue-specific gene expression regulation (Baek *et al.*, 2007; Bharti *et al.*, 2008; Davuluri *et al.*, 2006; Kakizawa *et al.*, 2007; Kimura *et al.*, 2006; Puomila *et al.*, 2007; Sehgal *et al.*, 2008). Moreover, the abnormal preferential usage of one promoter over the other for certain genes has been revealed to contribute to cancer development in various tissues (Bulun *et al.*, 2009; Davuluri *et al.*, 2006). Possible implications for preferential usage of the MLP-P1 promoter has raised questions with regards to whether or not the cellular environment indicative of breast cancer cells and tumours preferentially activates transcription of *mier1* isoforms at the MLP-P1 promoter over the MAEP-P2 promoter. As previously noted in sections 1.5.6 and 3.5, this possible phenomenon is the basis for a promising mechanism serving to explain MIER1 $\alpha$ 's aberrant subcellular localization in invasive breast tumours and cancer cells.

***The MLP-P1 promoter is consistently activated at a higher level than MAEP-P2 in all cell lines examined, the extent of which is highest in ER+ breast cancer cell lines.***

Results from section 3.4.1 demonstrate that MAEP-P2 RLU levels varied across the examined cell lines (Fig 3.11). Moreover, these reporter activity levels did not correlate specifically with cell line ER status as the two ER- cell lines produced considerably distinct RLU levels, an observation consistent with the MLP-P1 promoter results. It is evident from Figure 3.12 though, that the MLP-P1 promoter is

consistently activated at a higher level than MAEP-P2, thereby incessantly producing more luciferase activity than *mier1*'s second promoter in each cell line. This suggests that, in general, the MLP-P1 promoter is the prominently active promoter of *mier1*.

Intriguingly though, implications for a possible role of ER status in the regulation of *mier1* transcription surfaced when the luciferase activity data obtained from the aforementioned experiments between the two *mier1* promoters was compared (Fig 3.12 & Table 3.4). The degree to which MLP-P1 is activated over MAEP-P2 increases in ER+ breast cancer cell lines. This striking pattern is clearly evident upon the calculation of the ratios between the RLU levels stemming from the maximum activity- producing luciferase reporter gene deletion construct of each promoter [MLP-P1 (-185) pGL3 for MLP-P1 and MAEP-P2 (-312) pGL3 for MAEP-P2] as outlined in Table 3.4. This suggests that in ER+ breast cancers, which already contain heightened ER signalling, MLP-P1 would become even more activated over the MAEP-P2 promoter. A key potential repercussion of this preferential activation is discussed in the following section.

Furthermore, it is interesting that this phenomenon is prominent in ER+ breast cancers thereby suggesting that perhaps ER signalling may play an accessory role in dictating the degree to which MLP-P1 is activated over MAEP-P2. There are various ways to investigate this point. A promoter analysis approach similar to the approach used in this study could be employed in order to assay the effect of either inducing or abrogating ER signalling in these cells, such as with either estrogen treatment to activate estrogen signalling, and estrogen inhibitor treatment or si-RNA technology to knock down ER expression in order to impede ER signalling. This would then allow



one to monitor the promoter activity ratios between MLP-P1 and MAEP-P2 under these various conditions that either induce or disrupt ER signalling.

Intriguingly, previous *in silico* analysis of the MAEP-P2 promoter revealed that a ½ ERE resides -256bp from the MAEP-P2 TSS. Moreover, numerous studies have found that Sp1, a TF that was previously shown to regulate transcription at MAEP-P2 specifically, can function in concert with ER and its cognate ½ ERE sites to induce transcription of downstream target genes (Kim *et al.*, 2005; Porter *et al.*, 1997). Therefore one would postulate that the MAEP-P2 promoter would be more active than MLP-P1 (which does not harbour any such ½ or full EREs) in the presence of ER signalling. This study showed the exact opposite; thereby suggesting that the previously described MLP-P1 TFBS, or any possible non-characterized MLP-P1 TFBS, must be more conducive to attracting the transcription machinery at this promoter than the MAEP-P2 specific TFBS are for the MAEP-P2 promoter in the examined cell lines.

Finally, a key goal to studying promoter activity levels is to investigate how transcription is regulated at a particular gene of interest. With the finding that MLP-P1 is the predominantly activated promoter of *mier1*, the next pivotal step would be to examine whether or not this preferential activation affects the level and type of endogenous *mier1* transcripts in a particular cell line. It would be interesting to see whether or not ER+ breast cancer cells generate a higher percentage of MLP-P1 specific vs. MAEP-P2 specific transcripts than ER- breast cancer cell lines. Furthermore, transcript type and level could be investigated in the absence or presence

of estrogen treatment or estrogen signalling inhibition in these various breast cancer cell lines.

These types of analyses can be performed by first extracting the total RNA from a number of the examined breast cancer cell lines, either directly or following treatment with either estrogen or estrogen inhibitors (specifically in the case of ER+ breast cancer cells) and then performing reverse transcription (RT)-PCR to obtain sufficient cell line specific cDNA. This cDNA could then be run through various PCR reactions using primers designed specifically to anneal to each of the promoter-specific transcripts (for reference to all possible *mier1* transcripts, refer to Figure 1.6). A very important consideration with this method, however, is primer design. Each primer must be engineered so that they each anneal to the cDNA with equivalent efficiency thereby ensuring that any differences in the final observed PCR products are representative of the extracted RNA and not a function of differential primer annealing. Therefore, the stringent design of the transcript specific primers is very important, and should take into consideration any possible secondary structure formed by each primer, the annealing and denaturing temperatures of the primers, primer nucleotide content, and the sizes of the amplicons produced by each transcript specific pair. Following RT-PCR and PCR results, qPCR should then be performed to further verify and confirm the result trends, using new and carefully designed primers for each promoter specific transcript as corresponding to qPCR primer requirements.

Other methodological approaches could be employed for this transcript analysis, however, these also intrinsically contain certain caveats. For example, one could perform a northern blot using a probe that binds to the common region of all



transcripts. This northern blott analysis would then directly detect the extracted RNA and the differences in the sizes of the transcripts could then be use to distinguish a particular promoter specific transcript. However, as previously illustrated by Paterno *et al.* 2002, this assay would not be able to distinguish between all of the different *mier1* transcripts as the 5' transcript ends produced by either MLP-P1 and MAEP-P2 are not sufficiently different enough to resolve on a gel. Finally, an RNase protection assay could be performed with probes specific for each *mier1* transcript. This would; however, inherently require the careful and assiduous planning and design of efficiently annealing probes, not unlike the primer design needed for the previously proposed RT-PCR method.

#### **4.7 Implications for Preferential Transcriptional Activation at MLP-P1 and the Effect of MLP-P1 Specific Exon 3A Inclusion on the Subcellular Localization of MIER1 $\alpha$ .**

As described in section 1.5.6, transcriptional activation at MLP-P1 can potentially result in the incorporation of the alternate exon, exon 3A, in the nascent MLP-P1 specific transcript. Exon 3A specifies a 74bp nucleotide sequence that translates into the following amino acid sequence: MFMFNWFTDCLWTLFLSNYQ. The alternate N-terminal end resulting from exon 3A inclusion harbours many hydrophobic amino acids that, as predicted by *in silico* algorithms, could possibly function as a transmembrane domain, which could serve to anchor the resulting protein isoform to the plasma membrane. Additionally, there is also a putative nuclear export sequence (NES) that contains the consensus L(X2-3) L(X2-3) LXL (whereby

“L” represents leucine or any other large hydrophobic residue, namely valine, isoleucine, phenylalanine, or methionine and “X” represents an amino acid) (Dong *et al.*, 2009; Kosugi *et al.*, 2008; La Cour *et al.*, 2003; Rosas-Acosta & Wilson, 2008). In the amino acid sequence encoded by exon 3A this putative NES resides between the 7<sup>th</sup> and 16<sup>th</sup> amino acids and constitutes the following sequence: ETDCLWTLFL.

These two aforementioned putative characteristics of the N-terminal domain encoded by exon 3A have significant implications for the regulation of MIER1 isoform function. It is currently hypothesized that in order for MIER1 isoforms to fully exert their general regulatory role in the cell, they should be present in the nucleus. Therefore any action that would evoke a change in subcellular localization causing MIER1 to be removed from the nucleus, such as plasma membrane anchorage or nuclear export, could possibly abrogate or change MIER1's transcriptional regulatory functions. Furthermore, as discussed in section 1.5.5, with regards to MIER1 and breast cancer development, MIER1 $\alpha$  has shown to undergo a subcellular localization switch to the cytoplasm as breast carcinomas progress to a more invasive state. The inclusion of exon 3A, with either its putative transmembrane or nuclear export domains, may serve as a possible explanation as to why this isoform is no longer localized to the nucleus as breast cells become more malignant.

***The 74bp insert specified by exon 3A functions to localize MIER1 $\alpha$  to the cytoplasm in MCF7 cells***

Results outlined in section 3.5.1 following transient transfection of MCF7 cells with MYC-tagged *mier1* isoform expression vectors showed that the 74bp insert specified by exon 3A does in fact function to localize MIER1 $\alpha$  to the cytoplasm (Figures 3.15 & 3.16). This finding has significant implications, not only for the importance of regulation at MLP-P1, but for further describing the state and location of MIER1 $\alpha$  in various stages of breast cancer. These results suggest that if some event or cue from the cellular environment throughout breast cancer progression causes increased transcriptional activation at MLP-P1, and thus an increased production of transcripts harbouring exon 3A, then the protein isoforms subsequently produced could become localized in the cytoplasm.

An important future direction to this effect though, would be to analyze the actual transcript profile in breast cancer cells at different levels of breast cancer progression in order to determine whether there is a switch to the production of transcripts containing exon 3A. These analyses could be performed like the previously proposed transcript analysis, but instead of focusing on discriminating between the promoter specific isoforms, one would have to further distinguish between the MLP-P1 specific isoforms containing and lacking exon 3A. This would be possible by either designing primers that specifically span this particular exon, or by simply using amplicon size as a marker for exon 3A incorporation. For example, all MLP-P1 transcripts could be detected by using a specific 5' primer in the MLP-P1 UTR. Transcripts containing exon 3A, and therefore the PCR products generated from such



transcripts following RT-PCR and subsequent PCR reactions, would not run as far down an agarose gel as the transcripts lacking exon 3A as these PCR products would contain 74 extra nucleotides.

Further characterization of the subcellular localization, as well as the level of MLP-P1 specific *mier1* isoforms at various stages of breast cancer, could be investigated using immunohistochemical approaches employing tissue microarrays constituting various breast cancer tumour samples varying in grade and/ or degree of neoplasia (such as hyperplasia vs. carcinoma *in situ*). Antibodies specific for MIER1 $\alpha$  isoforms harbouring the amino acid sequence encoded by the exon 3A-specified 74bp insert could be developed in order to stain these tissues. One could then directly observe the subcellular localization of such isoforms in these various tissues. Moreover, co-staining of MIER1 $\alpha$  isoforms that originate from transcripts either containing or lacking the 74bp insert with antibodies specific for each separate isoform could be performed and analyzed using confocal microscopy. This technique would be extremely beneficial, as it would allow the investigator to compare the location of both of these MIER1 $\alpha$  isoforms in the same cell, at the same time, and in the same tissue. Then, such as with the previously proposed immunohistochemical analysis, confocal analysis comparing the isoforms could be utilized on distinct tissues of varying cancer grade, even in invasive breast cancer.

McCarthy *et al.* 2008 is currently the only study to have directly investigated MIER1 $\alpha$  in invasive breast cancer. As described in section 1.5.5, that particular study revealed that MIER1 $\alpha$  was predominantly localized in the cytoplasm as breast cancer cells progressed towards an invasive state. It would be interesting to examine whether

or not the MIER1 $\alpha$  isoforms that are found in the cytoplasm contain the amino acid sequences encoded by the 74bp insert specified by exon 3A. This would specifically be revealed following the immunohistochemical and confocal analysis described earlier. If these isoforms were discovered to contain such amino acids, then a role for the MLP-P1 promoter in invasive breast cancer progression; and as such, the mechanism of how these specific isoforms are getting localized to the cytoplasm in this particular stage of breast cancer would be further revealed.

*LMB treatment further supports the hypothesis that the 74bp insert does contain a functional NES which is responsible for the subcellular targeting of MIER1 $\alpha$  74bp-containing isoforms to the cytoplasm in MCF7 cells.*

The ability of LMB to efficiently inhibit nuclear export has greatly advanced the current knowledge surrounding nucleocytoplasmic trafficking (Kosugi *et al.*, 2008; La Cour *et al.* 2003). LMB treatment has furthermore been relentlessly utilized to confirm the functionality of putative leucine rich nuclear export signals (Amazit *et al.*, 2003). Researchers are able to use LMB treatment to help specifically pinpoint whether or not a putative NES can function as a true NES, as LMB treatment abrogates the functioning of a key exportin involved in the prime nuclear export signalling transduction pathway (Dong *et al.* 2009, Kudo *et al.* 1998; Kudo *et al.*, 1999). In theory, if a molecule harbouring the putative NES in question was found to localize to the cytoplasm, then LMB treatment would further support the functionality of the putative NES as a true NES by causing a decrease of cytoplasmic accumulation.



This exact phenomenon was exemplified in this study. LMB treatment of MCF7 cells caused a decrease in cytoplasmic accumulation of the *mier1* isoform harbouring the putative NES (MIER1 3A alpha) and reciprocally induced a higher percentage of nuclear and whole cell staining for this particular isoform (Fig 3.17). This finding fundamentally supports the hypothesis that the MLP-P1 driven alternate incorporation of exon 3A results in the expression of a NES which does contain nuclear export function via the CRM1 exportin-dependent nuclear exportation pathway in this ER+ breast carcinoma cell line.

*The subcellular localization patterns of the MIER1 $\alpha$  isoforms in ER- MDA MB 231 cells and the non-cancerous ER- HEK 293 cells did not recapitulate the subcellular localization patterns observed in the ER+ MCF7 cell line*

Interestingly, the subcellular localization results of the MIER1 $\alpha$  isoforms in ER- MDA MB 231 breast cancer cells and ER- HEK 293 cells did not parallel that which was found in the ER+ MCF7 breast cancer cells. In both cell lines, whole cell staining was the most prominent outcome following immunocytochemical staining for the different MIER1 $\alpha$  isoforms. This suggests that the MIER1 3A alpha isoform containing the amino acids encoded by the exon 3A-specified 74bp insert does not induce exclusive cytoplasmic subcellular localization of these particular proteins in these cells. Moreover MIER1 $\alpha$  isoforms are not predominantly localized in the nucleus in these cell lines, as is the case in MCF7 cells.

This finding raises a lot of pertinent questions, the most evident being: what is causing this substantial localization pattern difference across these cell lines? The most obvious feature that differentiates MDA MB 231 cells and HEK 293 cells from



MCF7 cells is ER signalling thereby suggesting that the ER may play a role in MIER1 $\alpha$  isoform localization. What is also interesting though, is a further question of how the MIER1 $\alpha$  isoform can localize to the nucleus in MCF7 cells in the first place, as MIER1 $\alpha$  isoforms do not have intrinsic nuclear localization signals (section 1.5.1, 1.5.3, 1.5.4). Ding *et al*, 2003 first addressed this issue by suggesting that MIER1 $\alpha$  may be cotransported to the nucleus through regulated interactions with another nuclear protein(s) through a type of piggyback mechanism (Ding *et al*, 2003). Such a mechanism has been revealed for many different proteins such as Hsp90, retinoblastoma (RB) gene product, mitogen activated protein kinase (MAPK), p53,  $\beta$  catenin (Davis *et al.*, 2006).

Could it be possible that this specific piggybacking protein is the ER? Such an occurrence may explain the observation of the exclusively nuclear staining in ER+ MCF7 cells but not in the ER- MDA MB 231 cells and the ER- HEK 293 cells. There are various experiments one could set up to investigate this. One option would be to perform the exact same type of immunocytochemistry experiment in MCF7 cells as described in this study, but to furthermore investigate whether knocking down ER expression in MCF7 cell lines, such as discussed in section 4.2, would abrogate MIER1 $\alpha$ 's nuclear localization. Furthermore, one could even transfect the ER into the ER- MDA MB 231 and HEK 293 cell lines to observe whether or not MIER1 $\alpha$  localizes and accumulates in the nucleus. In addition, another more direct option would be to coanalyze the subcellular localization of both the ER and MIER1 $\alpha$  isoforms in the same cell using confocal microscopy.

Confocal microscopy would not only be extremely useful for this particular proposed project, but would also serve to further confirm results from section 3.5 concerning the subcellular localization of the MIER1 $\alpha$  isoforms in MCF7 cells. Even though LMB treatment did adequately support a case for the putative NES as the functional component of the 74bp insert that induced MIER1 3A  $\alpha$  isoforms to move out of the nucleus, while also ruling out membrane anchorage as the functional mechanism, confocal microscopy could serve as an additional tool to further validate LMB treatment experiments. Using this specific technique one could effectively observe not only nuclear and/or cytoplasmic subcellular localization but more specifically plasma membrane anchorage accumulation much more clearly than would be possible using conventional immunocytochemical techniques.

Another important point with regards to the proposed ICC and the previously performed ICC experiments of this study is that future investigation needs to employ western blot analysis of the transiently transfected MIER1 $\alpha$  isoforms. This step is critical to these types of assays in order to confirm that the transfected MIER1 $\alpha$  isoforms are effectively being produced in the cells.

Finally, the incorporation of exon 3A may not solely be a function of preferential activation at MLP-P1. Another pivotal step involved in alternate exon incorporation is RNA splicing. As briefly described in section 1.2.2, like most cellular processes, RNA splicing is very complex and is intricately orchestrated by a variety of regulatory molecules and interconnecting pathways. In fact, some groups even attest that the spliceosome may represent the largest and most complex macromolecular structure in eukaryotic cells (Skotheim & Nees, 2007). In terms of

splicing profiles in tumourigenesis, there is increasing evidence that alternative and aberrant pre-mRNA splicing may play an important functional role in human cancers (Skotheim & Nees, 2007). Accumulating evidence suggests that the cellular splicing machinery is changed during oncogenic transformation and that a variety of candidate genes involved in RNA-processing are known to be up-regulated (Skotheim & Nees, 2007). Moreover, several tumour-specific splice variants have been characterized and it has been found that each of the classical “hallmarks of cancer” discussed previously are affected or sustained by at least a few known cancer associated splice variants (Skotheim & Nees, 2007).

Therefore, it is a wonder whether or not aberrantly functioning splicing machinery may be perpetuating exon 3A incorporation in MLP-P1 specific *mier1* transcripts. Regulation of *mier1*- specific splicing mechanisms has yet to be investigated, as research surrounding RNA splicing mechanisms in cancer development in general is still in its infancy (Pajares *et al.*, 2007). Future projects focusing on this particular research question though would not only reveal pertinent information surrounding the *mier1* regulatory network in cancer, but would further contribute to current knowledge surrounding the role of alternative splicing with regards to cancer progression.



#### 4.8 Final Summary of Conclusions & Implications

This study successfully characterized the nucleotide sequences that are involved in eliciting minimal and maximal activity at *mier1*'s MLP-P1 promoter. The results from this promoter activity mapping analysis facilitated the identification of candidate active *cis*-elements inherent to this promoter via *in silico* examination. The discovery of Sp1, CREB, C/EBP $\beta$ , USF, and c-REL TFBSs as candidate MLP-P1 promoter *cis*-elements suggests that MLP-P1 may be under the regulation of these TFs. Moreover, each of the aforementioned TFs, except USF, have been characterized as in playing a role in cancer, whether solely being dysregulated or serving directly to perpetuate the neoplastic phenotype. Therefore, an abnormal level of expression or activation of these particular TFs in cancer may serve to dysregulate *mier1* at MLP-P1 thereby ensuing abnormal production of MLP-P1 specific transcripts.

Further investigation of MLP-P1 specific regulation in a variety breast cancer cell lines and the non-cancerous HEK 293 cell line via promoter activity analysis reconfirmed the promoter activity of the established MLP-P1 minimal and maximal promoter sequences, as each cell line displayed similar promoter activity patterns. This finding implies that, in all of these cell lines, each section of the promoter is respectively being activated to the same extent, even though the overall levels of promoter activity differed across each cell type. The functional importance of the stretches of nucleotides involved in maximal and minimal activity at MLP-P1 were even further supported following conservation analysis across six different species

whereby four out of the six species analyzed displayed high levels of conservation and homology.

Analysis of *mier1* promoter activity across four different breast cancer cell lines showed that the MLP-P1 promoter is the predominant promoter in these cells. This observation further implicated breast cancer ER status in *mier1* transcriptional regulation as the ratio of MLP-P1 activation to MAEP-P2 activation was highest in ER+ breast carcinoma cell lines. The significance of this preferential transcriptional activation of MLP-P1 over MAEP-P2 lies in the alternate inclusion of *mier1*'s exon 3A, which is only possible following MLP-P1 specific *mier1* transcription.

MIER1 $\alpha$ 's subcellular localization changes from the nucleus to the cytoplasm as breast cancer cells progress to a more invasive state (McCarthy *et al.*, 2008). Exon 3A had recently been described as encoding for a 74bp sequence that harbours either a putative NES and/ or putative transmembrane domain, which suggests that the incorporation of this exon into *mier1* transcripts may produce *mier1* isoforms (referred to as MIER1 3A alpha in this study) that abnormally localize to the cytoplasm. ICC results specifically showed that MIER1 3A alpha isoforms do reside in the cytoplasm in ER+ MCF7 breast cancer cells. In contrast, MIER1 $\alpha$  isoforms produced from transcripts lacking exon 3A, termed MIER1 alpha, localized to the nucleus in these particular cells. This subcellular localization pattern changed following LMB treatment whereby cytoplasmic localization of MIER1 3A alpha decreased. This was furthermore followed by an increase in nuclear localization of MIER1 3A alpha, which suggested that this event was a function of NES mediated nuclear export. Interestingly though, when ER- MDA MB 231 cells and ER- HEK 293 cells were



analyzed, such subcellular localization patterns were not observed. Instead, both MIER1 $\alpha$  isoforms were found in both the cytoplasm and nucleus of these cells. These findings suggest that ER may play a role in the localization of these isoforms, as ER status is the prominent difference between these cell lines thereby necessitating further investigation.

This ICC data furthermore provided evidence that transcriptional activation of MLP-P1, and the subsequent generation of MIER1 $\alpha$  isoforms containing exon 3A, can result in sequestering of MIER1 $\alpha$  in the cytoplasm of ER+ breast cancer cells. Could this be a mechanism explaining what is happening to MIER1 $\alpha$  in invasive breast cancer? And if so, could cytoplasmic accumulation of this isoform directly enhance progression to invasive breast carcinoma? Is there a specific stage leading to the acquisition of invasive cancer cell characteristics whereby subcellular relocation of MIER1 $\alpha$  takes place? Could the change in subcellular localization of MIER1 $\alpha$  be used to indicate further transformation of a pre-existing benign carcinoma to invasive breast cancer? If so, is this correlation directly due to an abnormal function of an aberrantly localized MIER1 $\alpha$  and its interaction with ER $\alpha$ ? More inquiry and further investigation, like that which is discussed in the future directions suggested throughout this chapter, is undoubtedly needed to address these questions.

Fortunately, significant implications for such further investigation arise as pertinent clinical relevance may be revealed for this particular protein isoform in breast cancer patients. If MIER1 $\alpha$  cytoplasmic accumulation can be used as a prognostic factor indicating transformation to invasive cancer, clinicians may be able to more accurately diagnose and even possibly predict breast cancer stage, which would help selection of the most appropriate therapeutic regimes. Furthermore, if it is shown that there is a



functional correlation between the MLP-P1 MIER1 $\alpha$ -specific isoform and invasive breast cancer progression, *mier1* could be targeted using therapeutic modalities whereby transcriptional activation at MLP-P1 could be suppressed.

In conclusion, this study performed significant preliminary analyses that have revealed MLP-P1 as a very important component in the regulation of *mier1*. Future studies need to focus on investigating whether the differential promoter activity established in this study is reflected by transcript production. Furthermore, confocal microscopy and immunohistochemical analysis following the staining of MLP-P1 specific isoforms (MIER1  $\alpha$  and MIER1 3A  $\alpha$ ) in breast cancer tissues varying in tumour grade are needed to efficiently investigate the functional implications of exon 3A inclusion on invasive breast cancer progression. Finally, investigation is also required to delineate why subcellular localization patterns differ with respect to ER status. Results stemming from such further investigation and characterization of these important isoforms in different stages of breast cancer may thus reveal MIER1 $\alpha$  as a prognostic factor or even a therapeutic target in breast cancer development. Such a serendipitous discovery would then add MIER1 $\alpha$  to the repertoire of molecules vigorously trying to be collected in our efforts to make personalized medicine a reality in oncology and cancer treatment.

## References

- Aasland, R., Stewart, A. F., and T. Gibson (1996) The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIB. *Trends in Biochemical Sciences* 21, 87-88
- Abelev, G.I. and T. L. Eraiser (2008) On the Path to Understanding the Nature of Cancer. *Biochemistry (Moscow)* 73 (5), 487-497
- Agarwal, V. R., Bulun, S.E., Leitch, M., Rohrich, R., and R. Simpson (1996) Use of Alternative Promoters to Express the Aromatase Cytochrome P450 (CYP19) Gene in Breast Adipose Tissues of Cancer-Free and Breast Cancer Patients. *Journal of Clinical Endocrinology and Metabolism* 81 (11), 3843-3849
- Allison, M. (2008) Is personalized medicine finally arriving. *Nature Biotechnology* 26, 509-517
- Amazit, L., Alj, Y., Tyagi, R.K.T., Chauchereau, A., Loosfelt, H., Pichon, C., Pantel, J., Foulon-Guinhard, E., Leclerc, P., Milgro, E., and A. Guiochon-Mantel (2003) Subcellular Localization and Mechanisms of Nucleocytoplasmic Trafficking of Steroid Receptor Coactivator-1. *The Journal of Biological Chemistry* 278 (34), 32195-32203.
- Andres, J., Mai, K., Mohlig, M., Weickert, M. O., Bumke-Vogt, C., Diederich, S., Pfeiffer, A. F. H., Bahr, V., and J. Spranger (2007) Cell-type specific regulation of the human 11beta-hydroxysteroid dehydrogenase type 1 promoter. *Archives of Physiology and Biochemistry* 113(3), 110 – 115
- Angus, W.G.R., Larsen, M.C. and C. R. Jefcoate (1999) Expression of CYP1A1 and CYP1B1 depends on cell-specific factors in human breast cancer cell lines: role of estrogen receptor status. *Carcinogenesis* 20 (6), 947-955
- Aree, L., Yokoyama, N. N., and M. L. Waterman (2006) Diversity of LEF/TCF action in development and disease. *Oncogene* 25, 7492-7504
- Arnone, M.I., Dmochowski, I.J., and C. Gache (2004) Using Reporter Genes to Study cis-Regulatory Elements. *Methods in Cell Biology* 74, 621-652
- Badve, S., Turbin, D., Thorat, M. A., Morimiya, A. Nielsen, O.T., Perou, C. M., Dunn, S. , Huntsman, D. G., and H. Nakshatri (2007) FOXA1 Expression in Breast Cancer-Correlation with Luminal Subtype A and Survival. *Clinical Cancer Research* 13 (15), 4415-4421



- Baek, D. Davis, C., Ewing, B., Gordon, D., and P. Green (2007) Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. *Genome Research* 17, 145–155
- Bailey-Dell, K.J., Hassel, B., Doyle, A., and D. D. Ross (2001) Promoter characterization and genomic organization of the human breast cancer resistance protein (ATP-binding cassette transporter G2) gene. *Biochimica et Biophysica Acta* 1520, 234-241
- Bajic, V. B., Tan, S. T., Christoffels, A., Schonbach, C., Lipovich, L., Yang, L., Hofmann, O., Kruger, A., Hide, W., Kai, C., Kawai, J. Hume, D.A., Carninci, P., and Y. Hayashizaki (2006) Mice and Men: Their Promoter Properties. *PLoS Genetics* 2 (4), 0616-0626
- Bannister, A.J. *et al.* (2002) Histone methylation: dynamic or static? *Cell* 109, 801–806
- Baudot, A., Real, F., Izarzugaza, J., and I. Valenci (2009) From cancer genomes to cancer models: bridging the gaps. *EMBO Reports* 10, 359-366
- Belguise, K., Guo, S., Yang, S., Rogers, A.E., Seldin, D.C., Sherr, D.H., and G. E. Sonenshein (2007) Green Tea Polyphenols Reverse Cooperation between c-Rel and CK2 that Induces the Aryl Hydrocarbon Receptor, Slug, and an Invasive Phenotype. *Cancer Research* 67(24), 11742–50
- Ben-Yosef, T., Yanuka, O., Halle, D., and N. Benvenist (1999) Involvement of Myc targets in c-myc and N-myc induced human tumors. *Oncogene* 17, 165 -171
- Benz, C. C., Fedele, V., Xu, F., Ylstra, B., Ginzinger, D., Yu, M., Moore, D., Hall, R. K., Wolff, D. J., Disis, M. L., Eppenberger-Castori, S., Eppenberger, U., Schittulli, F., Tommasi, S., Paradiso, A., Scott, G. K., and D. G. Albertson (2006) Altered promoter usage characterizes monoallelic transcription arising with ERBB2 amplification in human breast cancers. *Genes Chromosomes Cancer* 45 (11), 983-994
- Bharti, K., Liu, W., Csermely, T., Bertuzzi, S., and H. Arnheiter (2008) Alternative promoter use in eye development: the complex role and regulation of the transcription factor MITF. *Development* 135, 1169–1178.
- Black, A. R., Black, J. D., and J. Azizkhan-Clifford (2001) Sp1 and Kruppel-Like Factor Family of Transcription Factors in Cell Growth Regulation and Cancer. *Journal of Cellular Physiology* 188, 143-160
- Black, D. L. (2003) Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry* 72, 291-336



- Blackmore, T.M., Mercer, C. F., Paterno, G. D., and L. L. Gillespie (2008) The transcriptional cofactor MIER1-beta negatively regulates histone acetyltransferase activity of the CREB-binding protein. *BMC Research Notes* 2008, 1(68), 1-14
- Blair, S., Bogerd, H. P., Madore, S. J., and B. R. Cullen (1994) Mutational analysis of the transcription activation domain of RelA: identification of a highly synergistic minimal acidic activation module. *Molecular and Cellular Biology* 14 (11), 7226-7234
- Boeger, H., Bushnell, D. A., Davis, R., Griesenbeck, J., Yahli, L., Strattan, J. S., Westover, K. D., and R. D. Kornberg (2005) Structural basis of eukaryotic gene transcription. *FEBS Letters* 579, 899-903
- Boyer, L. A., Latek, R. R., and C. L. Peterson (2004) The SANT domain: a unique histone-tail-binding module? *Nature Reviews Molecular Cell Biology* 5, 1-7
- Britton, D.J., Hutcheson, I.R., Knowlden, J.M., Barrow, D., Giles, M., McClelland, R.A., Gee, J.M., and R. I. Nicholson (2006) Bidirectional cross talk between ERalpha and EGFR signalling pathways regulates tamoxifen-resistant growth. *Breast Cancer Research Treatment* 96:131-146.
- Brudno, M., Poliakov, A., Minovitsky, S., Ratnere, I., and I. Dubchak (2007) Multiple whole genome alignments and novel biomedical applications at the VISTA portal. *Nucleic Acids Research* doi:10.1093/nar/gkm279: 1-6
- Bulun, S. E., Lin, Z., Zhao, H. Lu, M., Amin, S., Reierstad, S., and D. Chen (2009) Regulation of Aromatase Expression in Breast Cancer Tissue. *Steroid Enzymes and Cancer: Annals of the New York Academy of Sciences* 1155, 121-131
- Butler, J. E. F., and J. T. Kadonaga (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes & Development* 16, 2583-2592
- Cao, X., Zeng, J., and H. Yan (2008) Structural property of regulatory elements in human promoters. *Physical Review E* 77, 041908-1- 041908-7
- Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science* 309,1559-1563
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A.M., Taylor, M. S., Engstro, P. G., Frith, M. C., Forrest, A. R. R., Alkema, W. B., Lam, S., Plessy, T. C., Kodzius, R., Ravasil, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, S., Suzuki, H., Grimmond, S.M., Wells, C. A., Orlando, V., Wahlestedt, C., Liu, E. T., Harbers, M., Kawai, J., Bajic, V. B., Hume, D. A. and Y. Hayashizaki, (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics* 38 (6), 626-635

- Carroll, J.S., and M. Brown (2006) Estrogen receptor target gene: an evolving concept. *Mol Endocrinol* 20, 1707 -14
- Carroll, J.S, Liu, X.S., Brodsky, A.S., *et al.* (2005) Chromosomewide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 122, 33 -43.
- Canadian Cancer Society's Steering Committee (2009): Canadian Cancer Statistics 2009. Toronto: Canadian Cancer Society ISSN 0835-2976
- Cancer Genome Atlas Research Network. (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455, 1061-1068.
- Carlezon Jr, W.A., Dunman, R. S., and E. J. Nestler (2005) The many faces of CREB. *TRENDS in Neurosciences* 8, doi:10.1016/j.tins.2005.06.005
- Chen, D., Reierstad, S., Lu, M., Lin, Z., Ishikawa, H., and Bulun, S.E. (2009) Regulation of breast cancer-associated aromatase promoters. *Cancer Letters* 273, 15-27
- Chen, G. G., Zeng, Q., and G. M. K. Tse (2008) Estrogen and Its Receptors in Cancer. *Medicinal Research Reviews* 28 (6), 954-974
- Chen N., Szentirmay M.N., Pawar S.A., Sirito M., Wang J., Wang Z., Zhai Q., Yang H.X., Peehl D.M., Ware J.L., and M. Sawadogo (2006) Tumor-suppression function of transcription factor USF2 in prostate carcinogenesis. *Oncogene* 25:579 -587
- Clark G.M. (1995) Prognostic and predictive factors for breast cancer. *Breast Cancer* 2, 79-89.
- Clontech Laboratories, Inc. (2005) Tet-Off<sup>®</sup> and Tet-On<sup>®</sup> Gene Expression Systems User Manual. Clontech
- Corre, S., and M. Galibert (2005) Upstream stimulating factors: highly versatile stress-responsive transcription factors. *Pigment Cell Res.* 18, 337-348
- Corre, S., and M. Galibert (2007) USF : un régulateur essentiel de la transcription. *M/S: médecine sciences* 22 (1), 62-67
- Crick, F. (1970) Central dogma of molecular biology. *Nature* 227, 561-563
- Croce, C.M. (2008) Oncogenes and Cancer. *New England Journal of Medicine* 358, 502-511



Darnell Jr., J. E. (2002) Transcription Factors as Targets for Cancer Therapy. *Nature Reviews Cancer* 2, 740-749

Dauvois, S., White, R. and M. G. Parker (1993) The anti-estrogen ICI 182,780 disrupts estrogen receptor nucleocytoplasmic shuttling. *J. Cell Sci.* 106, 1377–1388.  
Davis, J. R., Kakar, M., and C. S. Lim (2006) Controlling Protein Compartmentalization to Overcome Disease. *Pharmaceutical Research* 24 (1), 17-27

Davuluri, R. V., Suzuki, Y., Sugano, S., Plass, C., and T. H. M. Huang (2008) The functional consequences of alternative promoter use in mammalian genomes. *Trends in Genetics* 24 (4), 167- 177

Dehma, S. M. and K. B. Bonham (2004) Regulation of alternative SRC promoter usage in HepG2 hepatocellular carcinoma cells. *Gene* 337, 141-150

Deutschman, C.S. (2005) Transcription. *Critical Care Medicine* 33 (12), S400-S403

Dhakshinamoorthy, S., Long II, D. J., and A. K. Jaiswal (2000) Antioxidant regulation of genes encoding enzymes that detoxify xenobiotics and carcinogens. *Current Topics in Cellular Regulation* 36, 201–206

Ding, L., Getz, G., Wheeler, D.A., Mardis, E.R., McLellan, M.D., Cibulskis, K., Sougnez, C., Greulich, H., Muzny, D.M., Morgan, M.B., *et al.* (2008). Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 455, 1069–1075

Ding, Z. (2004) Characterization of the Novel Transcriptional Regulator Human Mesoderm Induction Early Response Gene 1 (hMI-ER1): Its Promoters, Interacting Proteins and Transcriptional Regulatory Functions. (PhD Thesis, Memorial University of Newfoundland, 2004)

Ding, Z., Gillespie, L.L, Mercer, F.C., and G. D. Paterno (2004). The SANT domain of human MI-ER1 interacts with Sp1 to interfere with GC box recognition and repress transcription from its own promoter. *Journal of Biological Chemistry* 279, 28009-28016

Ding, Z., Gillespie, L.L, and G. D. Paterno (2003). Human MI-ER1 alpha and beta function as transcriptional repressors by recruitment of histone deacetylase 1 to their conserved ELM2 Domain. *Molecular Cell Biology* 23,250-258

Dobrzycka, K.M., Townson, S.M., Jiang, S., and S. Oesterreich (2003) Estrogen receptor corepressors – a role in human breast cancer? *Endocrine-Related Cancer* 10 517–536



- Dong, D., Jewell, C. M., Bienstock, R. J., and J. A. Cidlowski (2006) Functional analysis of the LXXLL motifs of the human glucocorticoid receptor: Association with altered ligand affinity. *Journal of Steroid Biochemistry & Molecular Biology* 101, 106–117
- Dong, X., Biswas, A., Suel, K.E., Jackson, L.K., Martinez, R., Gu, H., and Y.M. Chook (2009) Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature* 458, 1136–1142
- Dorling, A., Todman, M. G., Korach, K. S., and A. E. Herbison (2003) Critical Role for Estrogen Receptor alpha in Negative Feedback Regulation of Gonadotropin-Releasing Hormone mRNA Expression in the Female Mouse. *Neuroendocrinology* 78, 204–209
- Duan, R., Ginsburg, E., and B.K. Vonderhaar (2008) Estrogen stimulates transcription from the human prolactin distal promoter through AP1 and estrogen responsive elements in T47D human breast cancer cells. *Molecular Cell Endocrinology* 281(1-2), 9–18
- Dubchak, I., Poliakov, A., Kislyuk, A., and M. Brudno (2009) Multiple whole-genome alignments without a reference organism. *Genome Research* 19, 682–689
- Dupasquier, S., and C. Quittau-Prévostel (2009) Une expression dérégulée dans les cancers : des causes potentielles multiples. *C. R. Biologies* 332, 1–14
- Edelman, E. J., Guinney, J., Chi, J., Febbo, P. G., and S. Mukherjee (2008) Modeling Cancer Progression via Pathway Dependencies. *PLoS Comput Biol* 4(2): e28. doi:10.1371/journal.pcbi.0040028
- Emerson, B. M. (2002) Specificity of Gene Regulation *Cell* 109, 267–270
- Fawell, S.E., White, R., Hoare, S., Sydenham, M., Page, M. and M. G. Parker (1990) Inhibition of estrogen receptor-DNA binding by the ‘pure’ antiestrogen ICI 164,384 appears to be mediated by impaired receptor dimerization. *Proc. Natl Acad. Sci. USA* 87, 6883–6887.
- Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and I. Dubchak (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Research* 32, W273–W279
- Frith, M.C., Valen, E., Krogh, A., Hayashizaki, Y., Carninci, P., and A. Sandelin (2008) A code for transcription initiation in mammalian genomes. *Genome Res* 18, 1–12
- Fry, C.J., and C. L. Peterson (2002) Unlocking the Gates to Gene Expression. *Science* 295, 1847–1848

- Fukazawa, H., Mizuno, S., and Uehara, Y. (1995) A Microplate Assay for Quantitation of Anchorage Independent Growth of Transformed Cells. *Analytical Biochemistry* 228, 83-90
- Gaildrat, P., Meller, M., Mukada, S., Humphries, A., Carter, D.A., Ganapathy, V., and D.C. Klein (2005) A Novel Pineal-specific Product of the Oligopeptide Transporter *PepT1* Gene-Circadian Expression Mediated by cAMP Activation of an Intronic Promoter. *The Journal of Biological Chemistry* 280 (17), 16851–16860
- Gaspar Pereira S and F. Oakley (2008) Nuclear factor- $\kappa$ B1: Regulation and function. *The International Journal of Biochemistry & Cell Biology* 40 , 1425–1430
- Ghosh, S., Lu, Y., and Y. Hu (2008) A Role of CREB in BRCA1 Constitutive Promoter Activity and Aromatase Basal Expression. *Int J Biomed Sci.* 4(4), 260–265.
- Ginsburg, G.S. and J. J. McCarthy (2001) Personalized medicine: revolutionizing drug discovery and patient care. *Trends in Biotechnology* 19, 491-496
- Glover, J. F., Irwin, J. T., and P. D. Darbre (1988) Interaction of Phenol Red with Estrogenic and Antiestrogenic Action on Growth of Human Breast Cancer Cells ZR-75-1 and T-47-D. *Cancer Research* 48, 3693-3697
- Gonzalez-Angulo, A.M., Morales-Vasquez, F., and G. N. Hortobagyi (2007) Overview of resistance to systemic therapy in patients with breast cancer. *Adv Exp Med Biol.* 608, 1-22
- Gorski, J. J., Kennedy, R. D., Hosey, A. M., and D. P. Harkin (2009) The Complex Relationship between BRCA1 and ER $\alpha$  in Hereditary Breast Cancer. *Clinical Cancer Research* 15(5), 1514-1518
- Graham, F. L., and Smiley (1977) Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5. *J. gen. Virol.* 36, 59-7z
- Green, A. R., Burney, C., Granger, C. J., Paish, C., El-Sheikh, S., Rakha, E. A., Powe, D. G., Macmillan, D., Ellis, I. O., and E. Stylianou (2008) The prognostic significance of steroid receptor co-regulators in breast cancer: co-repressor NCOR2/SMRT is an independent indicator of poor outcome. *Breast Cancer Research Treatment* 110, 427–437
- Grimm, S.L., and J. M. Rosen (2003) The Role of C/EBP $\beta$  in Mammary Gland Development and Breast Cancer. *Journal of Mammary Gland Biology and Neoplasia* 8 (2), 191-204
- Grizzi, F., and M. Chiriva-Internati (2006) Cancer: looking for simplicity and finding complexity. *Cancer Cell International* 6 (4), doi:10.1186/1475-2867-6-4



Grizzi, F., Di Ieva, A., Russo, C., Frezza, E. E., Cobos, E., Muzzio, P. C., and M. Chiriva-Internati (2006) Cancer initiation and progression: an unsimplifiable complexity. *Theoretical Biology and Medical Modelling* 3(37), doi:10.1186/1742-4682-3-37

Grose, R., and C. Dickson (2005) Fibroblast growth factor signaling in tumorigenesis. *Cytokine and Growth Factor Reviews* 16, 179- 186

Gruvberger, S., Ringner, M., Chen, Y., Panavally, S., Sall, L. H., Borg, A., Ferno, M., Peterson, C., and P.S. Meltzer (2001) Estrogen Receptor Status in Breast Cancer Is Associated with Remarkably Distinct Gene Expression Patterns. *Cancer Research* 61, 5979- 5984

Gururaj, A. E., Rayala, S. K., Vadlamudi R. K., and R. Kumar (2006) Novel Mechanisms of Resistance to Endocrine Therapy: Genomic and Nongenomic Considerations . *Clinical Cancer Research* 12, 1001s-1007s

Hackett A.J., *et al.* (1977) Two syngeneic cell lines from human breast tissue: the aneuploid mammary epithelial (Hs 578T) and the diploid myoepithelial (Hs 578Bst) cell lines. *J. Natl. Cancer Inst.* 58: 1795-1806

Hall, J.M., and D. P. McDonnell (2005) Coregulators in nuclear estrogen receptor action: from concept to therapeutic targeting. *Molecular Interventions* 5, 343-357

Hahn, W. C, and R. A. Weinberg (2002) Modelling the Molecular Circuitry of Cancer. *Nature Reviews Cancer* 2, 331-341

Hanahan, D.and R. A. Weinberg (2000) The Hallmarks of Cancer. *Cell* 100, 57-70

Hansen, C., Howlin, J., Tengholm, S., Dyachok, O., Vogel, W.F., Nairn, A.C., Greengard, P., and T. Andersson (2009) Wnt-5a-induced Phosphorylation of DARPP-32 Inhibits Breast Cancer Cell Migration in a CREB-dependent Manner. *The Journal of Biological Chemistry* 284 (40), 27533–27543,

Hebden, C., Smalt, R., Chambers, T., and M. D. Pondel (2000) Multiple Promoters Regulate Human Calcitonin Receptor Gene Expression. *Biochemical and Biophysical Research Communications* 272, 738–743

Heery, D. M., Hoare, S., Hussain, S., Parker, M. G., and H. Sheppard (2001) Core LXXLL Motif Sequences in CREB-binding Protein, SRC1, and RIP140 Define Affinity and Selectivity for Steroid and Retinoid Receptors. *The Journal of Biological Chemistry* 276 (9), 6695–6702

Heery, D. M., Kalkhoven, E., Hoare, S., and M. G. Parker (1997) A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* 387, 733-736



Heintzman, N. D., and B. Ren (2007) The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome. *Cellular and Molecular Life Sciences* 64, 386-400

Hu, S., Chen, Z., Gu, W., Chen, R., Cen, J., and J. Cen (2008) The transcriptional activity of WT1 gene promoter and enhancer in cell lines with diverse tissue origin. *International Journal of Hematology* 87, 498-506

Hu Y.F., Lau K.M., Ho J.M., and J. Russo (1998) Increased expression of estrogen receptor beta in chemically transformed human breast epithelial cells. *International Journal of Oncology* 12, 1226-1228.

Huang, Q., Guo, J., Ge, Q., Li-Ling, J., Chen, X., and F. Ma (2009) Comparative analysis of distinct non-coding characteristics potentially contributing to the divergence of human tissue-specific genes. *Genetica* 136, 127-134

Hughes, T.A. (2006) Regulation of gene expression by alternative untranslated regions. *Trends in Genetics* 22(3), 119-122

Hughes, T. A., and H. J. M. Brady (2005) Expression of axin 2 is regulated by the alternative 5' untranslated regions of mRNA. *Journal of Biological Chemistry* 280, 8581-8588

Hume, D. A. (2008) Our evolving knowledge of the transcriptional landscape. *Mamm Genome* 19, 663-666

Inoue, H., Tanabe, T., and K. Umesono (2000) Feedback Control of Cyclooxygenase-2 Expression through PPAR. *Journal of Biological Chemistry* 275(36), 28028-28032

Ip, M.M., and B. B. Asch (2000) *Methods in Mammary Gland Biology and Breast Cancer Research* New York: Kluwer.

Ismail, R.M., Lu, T. and M. Le Sawadogo (1999) Loss of USF transcriptional activity in breast cancer cell lines. *Oncogene* 18, 5582 - 5591

Jain, K.K. (2002) Personalized Medicine. *Current Opinion in Molecular Therapeutics* 4, 548-558

Jaiswal, A. K. (2004) Nrf2 signaling in coordinated activation of antioxidant gene expression. *Free Radical Biological Medicine* 36, 1199-1207

Jones, L. P., Tilli, M.T., Assefnia, S., Torre, K., Halama, E.D., Parrish, A., Rosen, E. M. and P.A. Furth (2008) Activation of estrogen signaling pathways collaborates with loss of Brca1 to promote development of ER $\alpha$ -negative and ER $\alpha$ -positive mammary preneoplasia and cancer. *Oncogene* 27, 794-802

Jones, S., Zhang, X., Parsons, D.W., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A., et al. (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321, 1801–1806.

Kadonaga, J. T. (2004) Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell* 116, 247–257.

Kaestner, K.H. (2000) The Hepatocyte Nuclear Factor 3 (HNF3 or FOXA) Family in Metabolism. *TEM* 11(7), 281–285

Kakizawa T, Nishio S, Triqueneaux G, Bertrand S, Rambaud J, and V. Laudet (2007) Two differentially active alternative promoters control the expression of the zebrafish orphan nuclear receptor gene Rev-erbalpha. *Journal of Molecular Endocrinology* 38, 555–568.

Kaku, H., and T. L. Rothstein (2009) Octamer binding protein 2 (Oct2) regulates PD-L2 gene expression in B-1 cells through lineage-specific activity of a unique, intronic promoter. *Genes Immun* doi: 10.1038/gene.2009.68

Kang, K. I., Devin, J., Cadepond, D., Jibard, N., Guiochon-Mantel, A., Baulieu, E. E., and M. G. Catelli (1994) *In vivo* functional protein-protein interaction: nuclear targeted hsp90 shifts cytoplasmic steroid receptor mutants into the nucleus. *The Proceedings of the National Academy of Science USA* 91, 340–344

Kaspar, J.W., Niture, S. K., and A. K. Jaiswal (2009) Nrf2:INrf2 (Keap1) signaling in oxidative stress. *Free Radical Biology & Medicine*, doi:10.1016/j.freeradbiomed.2009.07.035

Kern, F. G., McLeskey, S.W., Zhang, L., Kurebayashi, J., Liu, Y., Ding, I. Y. F., Kharbanda, S., Chen, D., Miller, D., Cullen, K., Paik, S., and R. B. Dickson (2004) Transfected MCF-7 cells as a model for breast cancer progression. *Breast Cancer Research and Treatment* 31, 153–165

Kimura, K. Wakamatsu, A., Suzuki, Y., Ota, T. Nishikawa, T., Yamashita, R., Yamamoto, J., Sekine, M., Tsuritani, K., Wakaguri, H., Ishii, S., Sugiyama, T., Saito, K., Isono, Y., Irie, R., Kushida, N., Yoneyma, T., Otsuka, R., Kanda, K., Yokoi, T., Kondo, H., Wagatsuma, M., Murakawa, K., Ishida, S., Ishibashi, T., Takahashi-Fujii, A., Tanase, T., Nagai, K., Kikuchi, H., Nakai, K., Isogai, T., and S. Sugano (2006) Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res* 16, 55–65

Kininis, M., and W. L. Kraus (2008) A global view of transcriptional regulation by nuclear receptors: gene expression, factor localization, and DNA sequence analysis. *Nuclear Receptor Signaling* 6, e005



- Kim, K., Barhourni, R., Burghardt, R., and S. Safe (2005) Analysis of Estrogen Receptor  $\alpha$ -Sp1 Interactions in Breast Cancer Cells by Fluorescence Resonance Energy Transfer. *Molecular Endocrinology* 19(4), 843–854
- Kleinjan, D. A. and L. A. (2008) Long-Range Gene Control and Genetic Disease. *Advances in Genetics* 61, 339–388
- Ko, E. R., Ko, D., Chen, C., and J. S. Lipsick (2008) A conserved acidic patch in the Myb domain is required for activation of an endogenous target gene and for chromatin binding. *Molecular Cancer* 7(77), 1–20
- Koenigsberger, C., Chicca II, J. J. Amoureux, M. Edelman, G. M., and F. S. Jones (2000) Differential regulation by multiple promoters of the gene encoding the neuron-restrictive silencer factor. *Proceedings of the National Academy of Sciences* 97 (5), 2291–2296
- Kosugi, S., Hasebe, M., Tomita, M., and H. Yanagawa (2008) Nuclear Export Signal Consensus Sequences Defined Using a Localization-Based Yeast Selection System *Traffic* 9, 2053–2062
- Kroemer, G., and J. Pouyssegur (2008). Tumor cell metabolism: cancer's Achilles' heel. *Cancer Cell* 13, 472–482.
- Kudo, N., Wolff, B., Sekimoto, T., Schreiner, E.P., Yoneda, Y., Yanagida, M., Horinouchi, S., and M.Yoshida (1998) Leptomycin B Inhibition of Signal-Mediated Nuclear Export by Direct Binding to CRM1. *Experimental Cell Research* 242, 540–547
- Kudo, N. *et al.* (1999) Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region. *Proc. Natl Acad. Sci. USA* 96, 9112–9117.
- La Cour, T., Gupta, R., Rapacki, K., Skriver, K., Poulsen, F. M., and S. Brunak (2003) NESbase version 1.0: a database of nuclear export Signals. *Nucleic Acids Research* 31 (1), 393–396
- Lakowski, B., Roelens, I., and S. Jacob (2006) CoREST-Like Complexes Regulate Chromatin Modification and Neuronal Gene Expression. *Journal of Molecular Neuroscience* 29, 227–240
- Landry, J., Mager, D. L., and B. T. Wilhelm (2003) Complex controls: the role of alternative promoters in mammalian genomes. *Trends in Genetics* 19(11), 640–648
- Laganie, J., Deblois, D., Lefebvre, C., Bataille, A. R., Robert, F., and V. Giguere (2005) Location analysis of estrogen receptor  $\alpha$  target promoters reveals that FOXA1 defines a domain of the estrogen response. *PNAS* 102 (33) 11651–11656
- Lee, C., Shih, Y., Wu, C., and Y. A. Chen (2009) Characterization of the 5' regulatory region of the human Glycine N-methyltransferase gene. *Gene* 443, 151–157



Lemon, B. and R. Tjian (2000) Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev.* 14, 2551–2569.

Levenson, A. S., and V. C. Jordan (1997) MCF-7: The First Hormone responsive Breast Cancer Cell Line. *Cancer Research* 57, 3071-3078

Levine, M., and R. Tjian (2003) Transcription regulation and animal diversity. *Nature* 424, 147-152

Li, Q., Chu, M., and J. Xu (2007) Tissue- and Nuclear Receptor-Specific Function of the C-Terminal LXXLL Motif of Coactivator NCoA6/AIB3 in Mice. *Molecular and Cellular Biology* 27 (23), 8073–8086

Li, T. W. H., Ting, J. T., Yokoyama, N. N., Bernstein, A., Van de Wetering, M., and M. Waterman (2009) Wnt Activation and Alternative Promoter Repression of LEF1 in Colon Cancer. *Molecular and Cellular Biology* 26 (14), 5284-5299

Liou, H. and C. Y. Hsia (2003) Distinctions between c-Rel and other NF-kB proteins in immunity and disease. *BioEssays* 25, 767–780

Logette, E., Wotawa, A., Solier, S., Desoche, L., Solary, E., and L. Corcos (2003) The human caspase-2 gene: alternative promoters, pre-mRNA splicing and AUG usage direct isoform-specific expression. *Oncogene* 22, 935–946

Loignon, M., Miao, W. Hu, L., Bier, A., Bismar, T. A., Scrivens, P. J. Mann, K., Basik, M., Bouchard, A., Fiset, P. O. Batist, Z., and G. Batist (2009) Cul3 overexpression depletes Nrf2 in breast cancer and is associated with sensitivity to carcinogens, to oxidative stress, and to chemotherapy. *Molecular Cancer Therapy* 8(8), 2432–40

Lonard, D.M., and B. W. O'Malley (2007) Nuclear receptor coregulators: judges, juries and executioners of cellular regulation. *Molecular Cell* 27, 691–700

Loots, G.G. (2008) Genomic Identification of Regulatory Elements by Evolutionary Sequence Comparison and Functional Analysis. *Advances in Genetics* 61 269-293

Lopez-Tarruella, S., and R. Schiff (2007) The Dynamics of Estrogen Receptor Status in Breast Cancer: Re-shaping the Paradigm. *Clinical Cancer Research* 13(23), 6921-6925

Lu, S., and M. C. Archer (2009) Sp1 coordinately regulates de novo lipogenesis and proliferation in cancer cells. *Int. J. Cancer* 126, 416–425

Luo, J. Solimini, N.L., and S. J. Elledge (2009) Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction. *Cell* 136, 823-833

Manc, M. E., Hu, G., Sangster-Guilty, N., Olshalsky, S. L., Hoops, K., Fitzgerald-Bocarsly, P., Pitha, P., Pinder, K., and B. Barnes (2005) Two Discrete Promoters Regulate the Alternatively Spliced Human Interferon Regulatory Factor-5 Isoforms. *The Journal of Biological Chemistry* 280 (22), 21078-21090

Marro, M.L., Peiro, C., Panayiotou, C.M., Baliga, R.S., Meurer, S., Schmidt, H.H.H.W., and A.J. Hobbs (2008) Characterization of the Human Soluble Guanylyl Cyclase Promoter Key Role for NF- $\kappa$ B (p50) and CCAAT-Binding Factors in Regulating Expression of the Nitric Oxide Receptor. *The Journal of Biological Chemistry* 283 (29), 20027–20036

Matlin, A. J., Clark, F., and C. W. J. Smith (2005) Understanding Alternative Splicing: Towards a Cellular Code. *Nature Reviews Molecular Cell Biology* 6, 386-398

Maston, G.A., Evans, S.K., and M.R. Green (2006) Transcriptional regulatory elements in the human genome. *Annual Review of Genomics and Human Genetics* 7, 29–59.

Mayr, B. and M. Montminy (2001) Transcription Regulation by the phosphorylation dependent factor CREB *Nature Reviews Molecular Biology* 2, 599-609

Melcher, K. (2000) The strength of acidic activation domains correlates with their affinity for both transcriptional and non-transcriptional proteins. *Journal of Molecular Biology* 301, 1097-1112

Mercer, C. F., Paterno, G. P., Butt, K., and L.L. Gillespie (2010) MIER1 $\alpha$  Functions as an Estrogen Receptor- $\alpha$  Co-Repressor. unpublished data

Miller, K. D., and G. W. Sledge, Jr. (2003) Exploiting the hallmarks of cancer: the future conquest of breast cancer. *European Journal of Cancer* 39 (12), 1668-1675

McCarthy, P. L., Mercer, C. F., Savicky, M. W. J., Carter, B. A., Paterno, G. D., and L. L. Gillespie (2008) Changes in subcellular localisation of MI-ER1 $\alpha$ , a novel oestrogen receptor- $\alpha$  interacting protein, is associated with breast cancer progression. *British Journal of Cancer* 99, 639-646

McKenna, N.J., and B. W. O'Malley (2002) Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* 108,465–474

Mees, C., Nemunaitis, J., and N. Senzer (2009) Transcription factors: their potential as targets for an individualized therapeutic approach to cancer. *Cancer Gene Therapy* 16, 103-112



- Meitar, D., Crawford, S. E., Rademaker, A. W., and S. L. Cohn (1996) Tumor Angiogenesis Correlates With Metastatic Disease, N-myc Amplification, and Poor Outcome in Human Neuroblastoma. *J Clin Oncol* 14, 405-414.
- Miyazaki, K., Inoue, S., Yamada, K., Watanabe, M., Liu, Q., Watanabe, T., Adachi, M. T., Tanaka, Y., and S. Kitajima (2009) Differential usage of alternate promoters of the human stress response gene ATF3 in stress response and cancer cells. *Nucleic Acids Research* doi:10.1093/nar/gkn1082
- Mizukami, Y., Nonomura, A., Takizawa, T., Noguchi, M., Michigishi, T., Nakamura, S., and T. Ishizaki (1995) N-myc protein expression in human breast carcinoma: prognostic implications. *Anticancer Res.* 15(6B), 2899-905
- Mori, K., Fujii, R., Kida, N., Ohta, M., and K. Hayashi (1988) Identification of a polypeptide secreted by human breast cancer cells (MCF-7) as the human estrogen-responsive gene (pS2) product. *Biochemical and Biophysical Research Communications* 155(1), 366-372
- Morse, R. H. (2007) Transcription Factor Access to Promoter Elements. *Journal of Cellular Biochemistry* 102, 560-570
- Naar, A.M., Lemon, B. D., and R. Tjian (2001) Transcriptional Coactivator Complexes. *Annual Review of Biochemistry* 70, 475-501
- Nakanishi, T., Bailey-Dell, K. J., Hassel, B. A., Shiozawa, K., Sullivan, D.M., Turner, J., and D. D. Ross (2006) Novel 5' untranslated region variants of BCRP mRNA are differentially expressed in drug-selected cancer cells and in normal human tissues: implications for drug resistance, tissuespecific expression, and alternative promoter usage. *Cancer Research* 66, 5007-5011
- Naugler, W.E., and M. Karin (2008) NF- $\kappa$ B and cancer —identifying targets and mechanisms. *Current Opinion in Genetics & Development* 18, 19-26
- Naylor, L.H.(1999) Reporter Gene Technology: The Future Looks Bright *Biochemical Pharmacology* 58,749-757
- North, S., Espanel, X., Bantignies, F., Viollet, B., Vallet, V., Jalinot, P., Brun, G., and G. Gillet (1999). Regulation of cdc2 gene expression by the upstream stimulatory factors (USFs). *Oncogene* 18, 1945-1955.
- O Barrera, L. and B. Ren (2006) The transcriptional regulatory code of eukaryotic cells —insights from genome-wide analysis of chromatin organization and transcription factor binding. *Current Opinion in Cell Biology* 18, 291-298



Oh, D.S., Troester, M.A., Usary, J., Hu, Z., He, X., Fan, C., Wu, J., Carey, L.A., and C. M. Perou (2006) Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. *Journal of Clinical Oncology* 24,1656–64

O'Malley, B. (2008) The Year in Basic Science: Nuclear Receptors and Coregulators. *Molecular Endocrinology* 22 (12): 2751–2758

Orphanides, G., and D. Reinberg (2002) A Unified Theory of Gene Expression. *Cell* 108, 439–451

Osborne, C. K. (1998) Steroid hormone receptors in breast cancer management. *Breast Cancer Research Treatment* 51, 227–238

Otto, T., Horn, S., Brockmann, M., Eilers, U., Schuttrumpf, L., Popov, N., Kenney, A. M., Schulte, J. H., Beijersbergen, B., Christiansen, H., Berwanger, B., and M. Eilers (2008) Stabilization of N-Myc Is a Critical Function of Aurora A in Human Neuroblastoma. *Cancer Cell* 15, 67–78

Overdevest, J., Theodorescu, D., and J. Lee (2009) Utilizing the Molecular Gateway: The Path to Personalized Cancer Management. *Clinical Chemistry* 55 (4), 689-697

Panning B. and D. J. Taatjes (2008) Transcriptional Regulation: It Takes a Village. *Molecular Cell* 31, 622 - 229

Pajares, M.J., Ezponda, T., Caten, R., Calvo, A., Pio, R., and L. M. Montuenga (2007) Alternative splicing: an emerging topic in molecular and clinical oncology. *The Lancet* 8, 349-357

Park, S., Kim, D., Kaneko, S., Szewczyk, K.M., Nicosia, S. V., Yu, H., Jove, R., and J. Q. Cheng (2005) Molecular Cloning and Characterization of the Human *AKT1* Promoter Uncovers Its Up-regulation by the Src/Stat3 Pathway. *The Journal of Biological Chemistry* 280 (47), 38932–38941

Parsons, D.W., Jones, S., Zhang, X., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Siu, I.M., Gallia, G.L., *et al.* (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science* 321, 1807–1812.

Paterno, G.D., Ding, Z. Lew, Y., Nash, G.W., Mercer, C.F., and L. L. Gillespie (2002) Genomic organization of the human mi-er1 gene and characterization of alternatively spliced isoforms: regulated use of a facultative intron determines subcellular localization. *Gene* 295, 79-85

Paterno, G.D., Li, Y., Luchmann, H.A., Ryan, P.J., and L.L. Gillespie (1997) cDNA cloning of a novel , developmentally regulated immediate early gene activated by fibroblast growth factor and encoding a nuclear protein. *Journal of Biological Chemistry* 10, 25591-25595

Paterno, G.D, Mercer, C., Chayter, J.L., Yang, X., Robb, J.D., and L. L. Gillespie (1998) Molecular cloning of human *erl* cDNA and its differential expression in breast tumors and tumor-derived cell lines. *Gene* 222, 77-82

Payne, S. J. L., Bowen, R. L., Jones, J. L., and C. A. Wells (2008) Predictive markers in breast cancer – the present. *Histopathology* 52, 82–90

Pedersen, I.S., Dervan, P., McGoldrick, A., Harrison, M., Ponchel, F., Speirs, V., Isaacs, J.D., Gorey, T., and A. McCann (2002) Promoter switch: a novel mechanism causing biallelic PEG1/MEST expression in invasive breast cancer. *Human Molecular Genetics* 11, 1449-1453

Pennacchio L.A., and E. M. Rubin (2001) Genomic Strategies to identify mammalian regulatory sequences. *Nature Reviews Genetics* 2, 100-109

Phan, J. H., Moffitt, R. A., Stokes, T. H., Liu, J., Young, N. A., Nie, S., and M. D. Wang (2009) Convergence of biomarkers, bioinformatics and nanotechnology for individualized cancer treatment. *Cell Press* 27 (6), 350-359

Pink, J. J., Bilimoria, M.M., Assikis, J. and Jordan, V.C. (1996) Irreversible loss of the oestrogen receptor in T47D breast cancer cell following prolonged oestrogen deprivation. *Br. J. Cancer* 74, 1227–1236.

Plevin, M. J., Mills, M. M., and M. Ikura (2005) The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *TRENDS in Biochemical Sciences* 30 (2), 66-69

Polyak, K., (2007) Breast cancer: origins and evolution. *Journal of Clinical Investigation* 117, 3155–3163

Porter, W., Saville, B., Hoivik, D., and S. Safe (1997) Functional Synergy between the Transcription Factor Sp1 and the Estrogen Receptor. *Molecular Endocrinology* 11, 1569–1580

Post, J., Gillespie, LL. and G. D. Paterno (2001). Nuclear localization signals in the *Xenopus* FGF embryonic early response 1 protein. *FEBS Lett.* 502, 41-45

Puomila K, Simell O, Huoponen K, and J. Mykkanen (2007) Two alternative promoters regulate the expression of lysinuric protein intolerance gene SLC7A7. *Molecular Genetic Metabolism* 90, 298–306

Ravi, R. and A. Bedi (2004) NF- $\kappa$ B in cancer—a friend turned foe. *Drug Resistance Updates* 7, 53–67

Ray D., Terao Y., Fuhrken P.G., Ma Z.Q., DeMayo F.J., Christov K., Heerema N.A.,



Franks R., Tsai S.Y., Papoutsakis E.T., and H. Kiyokawa (2007) Deregulated CDC25A expression promotes mammary tumorigenesis with genomic instability. *Cancer Res* 67, 984–91.

Renaud, S., Pugacheva, E. M., Delgado, D., Braunschweig, R., Abdullaev, Z., Loukinov, D., Benhattar, J., and V. Lobanenko (2007) Expression of the CTCF-paralogous cancer-testis gene, brother of the regulator of imprinted sites (BORIS), is regulated by three alternative promoters modulated by CpG methylation and by CTCF and p53 transcription factors. *Nucleic Acids Research* 35 (21), 7372-7388

Risili, A.K., Shao, Z., Baumann, R. G., Li, X., Sheikh, S.M., Rimimi, S., Bashirelahi, N., and J. A. Fontana (1995) Estradiol Regulation of the Human Retinole Acid Receptor a Gene in Human Breast Carcinoma Cells Is Mediated via an Imperfect Half-Palindromic Estrogen Response Element and Spl Motifs. *Cancer Research* 55. 4999-5006.

Rosas-Acosta, G., and V. G. Wilson (2008) Identification of a nuclear export signal sequence for bovine papillomavirus E1 protein. *Virology* 373, 149–162

Safe, S., and M. Abdelrahim (2005) Sp transcription factor family and its role in cancer. *European Journal of Cancer* 41, 2438–2448

Safe, S., and K. Kim (2004) Nuclear receptor-mediated transactivation through interaction with Sp proteins. *Prog Nucleic Acid Res Mol Biol.* 77, 1-36

Sakakibara, Y., Irie, T., Suzuki, Y., Yamashita, R., Wakaguri, H., Kanai, A., Chiba, J., Takagi, T., Mizushima-Sugano, J., Hashimoto, S., Nakai, K., and S. Sugano (2007) Intrinsic Promoter Activities of Primary DNA Sequences in the Human Genome. *DNA Research* 14, 71-77

Sasada, R., Kurokawa, T., Iwane, M., and K. Igarashi (1988) Transformation of mouse BALB/c 3T3 cells with human basic fibroblast growth factor cDNA. *Molecular Cell Biology* 8 (2), 588-594

Savkur, R. S., and T. P. Burris (2004) The coactivator LXXLL nuclear receptor recognition motif. *Journal of Peptide Research* 63, 207–212

Schiff, R., Massarweh, S.A., Shou, J., Bharwani, L., Arpino, G., Rimawi, M., and C. K. Osborne (2005) Advanced concepts in estrogen receptor biology and breast cancer endocrine resistance: implicated role of growth factor signaling and estrogen receptor coregulators. *Cancer Chemotherapy and Pharmacology* 10-20.

Schiff, R., Massarweh, S., Shou, J., and C. K. Osborne (2003) Breast Cancer Endocrine Resistance How Growth Factor Signaling and Estrogen Receptor Coregulators Modulate Response. *Clinical Cancer Research* 9, 447S-454s



Sehgal A, Hughes B.T., and P.J. Espenshade (2008) Oxygen-dependent, alternative promoter controls translation of *teo1* in fission yeast. *Nucleic Acids Research* 36, 2024–2031.

Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.L., Wu, L., Lindeman, G.J., and J. E. Visvader (2006) Generation of a functional mammary gland from a single stem cell. *Nature* 439, 84-88.

Shah, N., Couronne, O., Pennacchio, L.A., Brudno, M., Batzoglou, S., Bethel, E.W., Rubin, E.M., Hamann, B., and I. Dubchak (2004) Phylo-VISTA: interactive visualization of multiple DNA sequence alignments. *Bioinformatics* 20 (5), 636–643

Sherr, C. J. (2004) Principles of Tumor Suppression. *Cell* 116, 235-246

Singaraja, R. R., Bocher, V., James, E.R., Clee, S. M., Zhang, L., Leavitt, B. R., Tan, B., Brooks-Wilson, A., Kwok, A., Bissada, N., Yang, Y., Liu, G., Tafuri, S. R., Fievet, C., Wellington, C. L., Staels, B., and M. R. Hayden (2001) Human *ABCA1* BAC Transgenic Mice Show Increased High Density Lipoprotein Cholesterol and ApoAI-dependent Efflux Stimulated by an Internal Promoter Containing Liver X Receptor Response Elements in Intron 1. *The Journal of Biological Chemistry* 276 (36), 33969–33979

Singhal, H. Guo, L., Bradlow, H.L., Mittelman, A., and R. K. Tiwar (1999) Endocrine Characteristics of Human Breast Epithelial Cells, MCF-10F. *Hormone Research* 52, 171–177

Skotheim, R.I, and M. Nees (2007) Alternative splicing in cancer: Noise, functional, or systematic? *The International Journal of Biochemistry & Cell Biology* 39, 1432–1449

Smale S.T. and J. T. Kadonaga (2003) The RNA Polymerase II Core Promoter. *Annual Review of Biochemistry* 72, 449–79

Smith, C.L., and B. W. O'Malley (2004) Coregulator function: a key to understanding tissue specificity of selective receptor modulators. *Endocrine Review* 25, 45-71

Smith, L. (2008) Post-transcriptional regulation of gene expression by alternative 5'-untranslated regions in carcinogenesis. *Biochemical Society Transactions*. 36, 708-711

Solari, F., Bateman, A., and J. Ahringer (1999) The *Caenorhabditis elegans* genes *egl-27* and *egr-1* are similar to MTA1, a member of a chromatin regulatory complex, and are redundantly required for embryonic patterning. *Development* 126, 2483-2494

Sonenberg, N., and A. G. Hinnebusch (2009) Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* 136, 731-745

Speirs, V., and R. A. Walker (2007) New perspectives into the biological and clinical relevance of oestrogen receptors in the human breast. *The Journal of Pathology* 211, 499-506.

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H.I., and C. J. Eaves (2006) Purification and unique properties of mammary epithelial stem cells. *Nature* 439, 993-997.

Stoneley, M., and A. E. Willis (2003) Aberrant Regulation of Translation Initiation in Tumorigenesis. *Current Molecular Medicine* 3, 597- 603

Skotheim, R.I., and M. Nees (2007) Alternative splicing in cancer: Noise, functional, or systematic? *The International Journal of Biochemistry & Cell Biology* 39, 1432-1449

Szutorisz, H., Dillon, N., and L. Tora (2005) The role of enhancers as centres for general transcription factor recruitment. *Trends in BioChemical Sciences* 30 (11), 593-599

Thompson, A., Brennan, K., Cox, A., Gee, J., Harcourt, D., Harris, A., Harvie, M., Holen, I., Howell, A., Nicholson, R., Steel, M., and C. Streuli (2008) Evaluation of the current knowledge limitations in breast cancer research: a gap analysis. *Breast Cancer Research* 10 (R26), doi:10.1186/bcr1983

Thorne, L. B., Grant, A. L., Paterno, G. P., and L. L. Gillespie (2005) Cloning and characterization of the mouse ortholog of *mi-er1*. *DNA Sequence* 16 (3), 237- 240

Trinklein, N. D., Force Aldred, S. J., Saldanha, A. J., and R. M. Myers (2003) Identification and functional analysis of human transcriptional promoters. *Genome Research* 13, 308-312

Turner, J. D., Schote, A. B., Macedo, J. A., Laetitia, A. B., Pelascini, P. L. and C. P. Muller (2006) Tissue specific glucocorticoid receptor expression, a role for alternative first exon usage? *Biochemical Pharmacology* 72, 1529-1537

Valle-Rios, R., No-Lopez, G.P., Medina-Contreras, O., Canche-Pool, E., Recillas-Targa, F., Lopez-Bayghen, E., Zlotnik, A., and V. Ortiz-Navarrete (2009) Characterization of CRTAM gene promoter: AP-1 transcription factor control its expression in human T CD8 lymphocytes. *Molecular Immunology*  
doi:10.1016/j.molimm.2009.07.016

Venables, J. P. (2009) Unbalanced alternative splicing and its significance in cancer. *BioEssays* 28, 378-386



Villard, J. (2004) Transcription regulation and human diseases. Swiss MED Weekly 134, 571-579

Walker, W. H., Girardet, C., and J. F. Habener (1996) Alternative exon splicing controls a translational switch from activator to repressor isoforms of transcription factor CREB during spermatogenesis. Journal of Biological Chemistry 271, 20145–20150

Wang, L., Charroux, B., Kerridge, S, and C. Tsai (2008) Atrophin recruits HDAC1/2 and G9a to modify histone H3K9 and to determine cell fates. EMBO reports 9, 555–562.

Wang, Q., Yang, W., Uytingco, M.S., Christakos, S., and R. Wieder (2000) 1,25-Dihydroxyvitamin D3 and All-*trans*-Retinoic Acid Sensitize Breast Cancer Cells to Chemotherapy-induced Cell Death Cancer Research 60, 2040–2048,

Wang, X.B., Peng, W.Q, Yi, Z.J., Zhu, S. L., and Q. H. Gan (2007) Expression and prognostic value of transcriptional factor sp1 in breast cancer. Ai Zheng 26(9) , 996-1000

Wei W.Z., Pauley R., Licklyter D., Soule H., Shi W.P., Calaf G., Russo J., and R. F. Jones (1998) Neoplastic progression of breast epithelial cells: A molecular analysis. Br J Cancer 78:198–204.

Werner, T. (2003) The state of the art of mammalian promoter recognition. Briefings in Bioinformatics 4 (1), 22-30

Whetstone, J. R., Flateley, R.M., and L. H. Matherly (2002) The human reduced folate carrier gene is ubiquitously and differentially expressed in normal human tissues: identification of seven non-coding exons and characterization of a novel promoter. Biochem Journal 367, 629-640

Xin D, Hu, L., and X. Kong (2008) Alternative Promoters Influence Alternative Splicing at the Genomic Level. PLoS One 3 ( 6), e2377

Yager, J. D., and N. E. Davidson (2006) Estrogen Carcinogenesis in Breast Cancer. The New England Journal of Medicine 354, 270-82.

Zacksenhaus, E., Jiang, Z., Hei, Y. J., Phillips, R. A., and B. L. Gallie (1999) Nuclear localization conferred by the pocket domain of the retinoblastoma gene product. Biochim. Biophys. Acta. 1451, 288-296

Zahnow, C. A. (2009) CCAAT/enhancer-binding protein b: its role in breast cancer and associations with receptor tyrosine kinases. Expert Reviews in molecular medicine 11 (e12), 1-29



Zhang T., Haws, P., and Q. Wu (2004) Multiple Variable First Exons: A Mechanism for Cell- and Tissue-Specific Gene Regulation. *Genome Research* 14, 79–89

Zheng, L., Annab, L. A., Afshari, C. A., Lee, W.H., and T. G. Boyer (2001) BRCA1 mediates ligand-independent transcriptional repression of the estrogen receptor. *The Proceedings of the National Academy of Sciences* 98, 9587-9592

**Appendix 1 – *mier1* MLP-P1 Promoter Region showing Location of 5' Starting Points for Insertion Sequences that were cloned into MLP-P1 Luciferase Reporter Gene Constructs**

```

(-1849/-3128)..... atgtttacat acacaaatact tagcatttgc ttacaattgc
(-1809/-3088)ctacagtatt cagtacagta acatgctatg  MLP-P1(-1708) pGL3 cagggttggtg gccatagaagc aacaaatgca
(-1749/-3028)ttttcatcat tatgcaaaaca ccatttagtg tacttacaca aacctagatg tatagcctac
(-1689/-2968)tatagcttag gtgtacagtc tactatacaa ttgtagccta qqtgtataqt aqcccataca
(-1629/-2908)tctacgtttg tgtaagtaca ctagaagata ttgcaaaat gctgaaattg gctaacaaag
(-1569/-2848)tattttctcag aagggtatcac catcgatgca tgactgtatt tagttatqqt agtqaggaac
(-1509/-2788)agataccagc aaaggatctg ggaaagggtga aacctatgaac caagacaaqa actaggaaga
(-1449/-2728)gtggtttctc agaaccctaaa taaagctttt caaagatcga gaagcgataa acccgaccga
(-1389/-2668)cactgcttca ttcaaaaaga atqaaactgt ggatcagcca ttgtatttgg ttgcattggt
(-1329/-2608)attgctggct ttcatagaag ctgttttgat ggagtggtag gatggacaaa aqcgtagta
(-1269/-2548)ccatattcaa gagagaatgc aaggttctct cccaagaact ctcaacaagt tccgctgaag
(-1209/-2488)aaagattagc tgagggtgag gagattqqag aggggaaqag aaaacaaatt acttagggaa
(-1149/-2428)ttttggggga acttagtggg cagcattacg qgcagcgcta agqaacnatt taaagtaaga
(-1089/-2368)caagtcaca  MLP-P1(-1077) pGL3 ctttcttaca gtcttggtca artgtgcat aaraacagaa
(-1029/-2308)tggtggaagc aggaattagt tttaaagtaa gatggtattg acgaacagac aafadaatac
(-969/-2248)ctctactaca gaggacagca atgttcggcc tgggtggtca cagaagctga qctctttqaa
(-909/-2188)cgctcgggca cagcatgaaa gcatcaagct tgagagactc cagqctttaa caaactaqaa
(-849/-2128)gtgctagaac gatggtteta aaccagggtct ttctcacac agcaagracca gcaatagaga
(-789/-2068)tttattttat cttatttttt agacagagtc  MLP-P1(-742) pGL3 tcgctctggt gctcagqctc qatqcaacg
(-729/-2008)gcacgatctc ggcttactgc aacctcggcc ttccgggttc aagggtattct cctgcctcag
(-669/-1948)ctcccaggt agctgggatt acaggcatgt gccaccaagt ccqgctaaat ttttgtattt
(-609/-1888)ttagtacaga cggggtttca ccattgttgc cagcctggac gntaacttct gacctcaggt

```

```

(-549/-1828) gatccacccc actcggcctc ccaaaagtccc gggattacag gtgtgaqena ccqcgcccag

(-489/-1768) ccnagtagag attcttttaa agactgcggt aggtcgataa aqqaqacatt qlqqcctagg

(-429/-1708) agcttcaagg tctacagact gaacccaqtl agggtagggc ccttttccaa atccaqaaat

(-369/-1648) qgtatctatt aatagaaggg ctctctcggg gtcccagggg gatqgaatcc qgttttttagc

(-309/-1588) qccggaactt acaaccggaa gttctgctgc ttgtgactgc ctqccgqaqa qactlgcaatg

(-249/-1528) gtggcgacca gctggggagt ggtqcaaccac cctttttttt ggccgcctct gaagtccctg

(-189/-1468) MLP-P1(-185) pGL3
tacccccagg ctctctcgtt agcggctcgg gccgaggtc cggaatgttt gccggggcgtc

(-129/-1408) MLP-P1(-91) pGL3
atqcgacagg tggagccctg gctcaacaag cggccgcgcg gttggctggc qgcacqaqgc

(-69/-1348) MLP-P1(-44) pGL3
cgaggaggaq ggccgagggc gaggggaggg cagaggggtg gtqgaactga aqqaagctcc

Putative TSS/ Exon 1A MLP-P1(+37) pGL3
(-9/-1288) ggaacgaacC TGAAGAAGG AGGCGGGCGG CCGGGGCCTC AGGCCCTCC CAGGCTCTGA

(+52/-1228) GTCTCCCGGC TGCAGGCGGA TGGATGGGGC TTCTTCAGGC GGTGGCGGCA GCAGCGAAGG

(+112/-1168) TGGCGGCGGC AGCAGCGGCA GCGGCT.....1046bp Intron.....

Exon 2A
(+1181/-95) ...ATGGTCT GGTCTGCTCGA TTCTCCCACT GCCTGGCTGA GTTTCGGACG TGGTTAAGAA

(+1241/-38) CCAACTGGTT GAGGTTCAAT GCAGACAAGA CGGATGTCAT GGTG

```

- Nucleotide counts include the number of base pairs from the putative TSS (first number) and the ATG translation start site (second number). The 1046bp intron is also included in the count. The 5' starting position of each MLP-P1 luciferase reporter gene deletion construct is indicated. The capital letters indicate UTR sequences originating from either exon 1A or 2A as marked. The TSS is also denoted.



## Appendix 2- *mier1* MLP-P1 Region Intron Sequence

```
(+138/-1178) .....gta agtgcagcct ccacaagcca
(+198/-1118) tctctcccct tctattccag tttaggatga ggggctcctg aggtgtcctc agtccccttt
(+258/-1058) ctctccttcc cctcccccac ggcagtgtag cgcccggaag acccttaact tccggggagg
(+318/-998) ggctgccgca gaacgcgcct ggcttgctct ccgccgctg ccaaaggcgc atgcgcagct
(+378/-938) tctcctctct gcccatcgac tgcctcccag cgccgccttt ttgccttcgc ggtggtggcg
(+438/-878) ccgcgctggg aatccgctgc ggagtgaagt cgccctcgcc cgttcgcttc ggtccctgat
(+498/-818) cccagtcggg gtggggctaa gctgaccacc ctgctgtggc tccgcctctt tctcctgtat
(+558/-758) ttccctcact tgtgtcccat ccccgaggag ctctcgcttg cactgcagcc ttlatggtga
(+618/-698) gcagcgccct gggccaactt gccctttctg gatggagtca gggagggagg agagactcga
(+678/-638) atagggattt actgtaagga aacgaggtcc tcttagcgcg tgggaggatc ccttgggaca
(+738/-578) gatgcccgcc gctctttact cttgttcttt cagcttgtag tggctaggtc cctgacctgg
(+798/-518) agcctatctg tggccctgc atgctgtgta accttgagcg agtcactatc cctctctggg
(+858/-458) cttcagtttt ccctatttaa agcttccctc aacgctagaa ttctgcattt ctgcaqaaga
(+918/-398) acttatgtaa ttacaagatt gacttggtta ttgaggccgc ttcaccaaga tgacccggag
(+978/-338) ttgactgaac cttcttcaact tatgggtgga atgaggcgac ctattttatt ggtcgctttt
(+1038/-278) attcctttat cattagttac aaattgcatg acttaaggct cattaaatgt ttttctaaaa
(+1098/-218) ttgtcatggg atccttgtcc tgattcccga ccatcttttt aaaaatgcga aacaggggtg
(+1158/-158) atggtgagct tgtatcatcg tttctgtctc cttgtactg aggctctctt tcccttgatc
(+1218/-98) cagATGGTGT GGTGCTCGA TTCTCCAGT GCCTGGCTGA GTTTCGACG TGGTTAAGAA
(+1279/-38) CCAACTGGTT GAGGTTCAAT GCAGACAAGA CGGATGTAT GTG
```

- Capital letters denote Exon 2A. The ATG translation start site of MLP-P1 is indicated by the red box. Nucleotide counts include the number of base pairs from the putative TSS (first number) and the ATG translation start site (second number).

### Appendix 3- *mier1* MAEP-P2 Promoter Region & Location of Primers used to Engineer Luciferase Reporter Gene Constructs

<b>MAEP-P2 (-1316) pGL3</b>						
(-1316/-1465)	cagactgtct	gtggactctt	ttcctgtcaa	attaccaggt	aatgtctttt	cacttagcag
(-1256/-1405)	ctacttcata	tatctgtaaa	tgcagtttat	tcattgcacag	attattattg	gacataacag
(-1196/-1345)	tttcaaatta	acattqaatc	tttgaaata	ggaaaatgct	tgataqaqtc	tqaattttta
(-1136/-1285)	agggaaatata	aataatttaa	tgcctccccc	tatattgcat	tataaqttaa	gtattaaactt
(-1076/-1225)	taagaaaaat	tgtatatgta	cacatttatg	caatgtaqta	ttggttcaca	ttatttqgaa
(-1016/-1165)	ggaatgaata	attctcgtat	ttaaatggtt	gatttttggt	gcacttttg	aaactgcatt
(-956/-1105)	aaqaggaacc	cttgaagatt	agqaaaaaaa	tcccagtcct	ttqaaataaa	qactttcaaa
(-896/-1045)	ctggtttagt	aaaataatgtt	agtgactgtt	ttaatgtcac	ctttgtqta	ttcagtatgt
(-836/-985)	gccttagtat	aatggaagt	aaacatqta	gttactttga	aatcatttgt	tctggccatc
(-776/-925)	atcagcattg	gcacacggc	caattqqcag	ctgtaggcca	atacattaaa	ataaacttca
(-716/-865)	agttacaggc	atttcaggca	ttgcaaaata	gtaggttctc	cctccatgct	actaaataga
(-656/-805)	tgatatccat	aaaacagtca	attttcagta	ttttaaattt	tgtctagtgt	tgcaaaattc
(-596/-745)	agtacataac	tgtgttgcta	ccggcatgac	agaagtgtgt	gagaaaaagc	aatgacatg
(-536/-685)	aggatggcgg	actgcctatc	ataacccctt	ccatttaattt	ttgctctqac	ctctctagag
(-476/-625)	atttqgggat	atagaatttt	acatttctct	tcgggaatct	actgatqqaq	aaatatttgt
(-416/-565)	caaggcataa	ccgctttcaa	agcacagttc	ttgccgcctt	tgcatcacag	caacatccgt
(-356/-505)	ttttcagcaa	atgcatttca	aaaacqacct	acatgtaaa	tat-tacgtat	ttttctctctg
<b>MAEP-P2 (-312) pGL3</b>						
(-296/-445)	ctgtgtcaat	gctggatgtg	actcctttgg	cactgttctt	tgacctctct	gatergtaga
(-236/-385)	caaccccttc	cgctggaggt	ggcggaacag	ctggagccag	cgaagcgcct	cgcgcggttg
(-176/-325)	cccacctctt	cccacaccca	ctttgaactc	gccccctccg	ctcttcccgq	qqagggctgg



**MAEP-P2 (-68) pGL3**

(-116/-265) ccgcggggccc gcgcgcgcgc cctgctccg gcgcgtgctc gctggtcttt tccctccagt

(-56/-205) ccagcccagc cggggcgccg cgagggggcg gagtggggtg tgggggcgc gctcgc**TSS**CGG

**MAEP-P2 (+28) pGL3**

(+5/-145) CTCCTGCGCG TTCCCGCCGA GGCAGTGGCG GCGGGAGCGG CAGAGACGGC AGCGGCCGGA

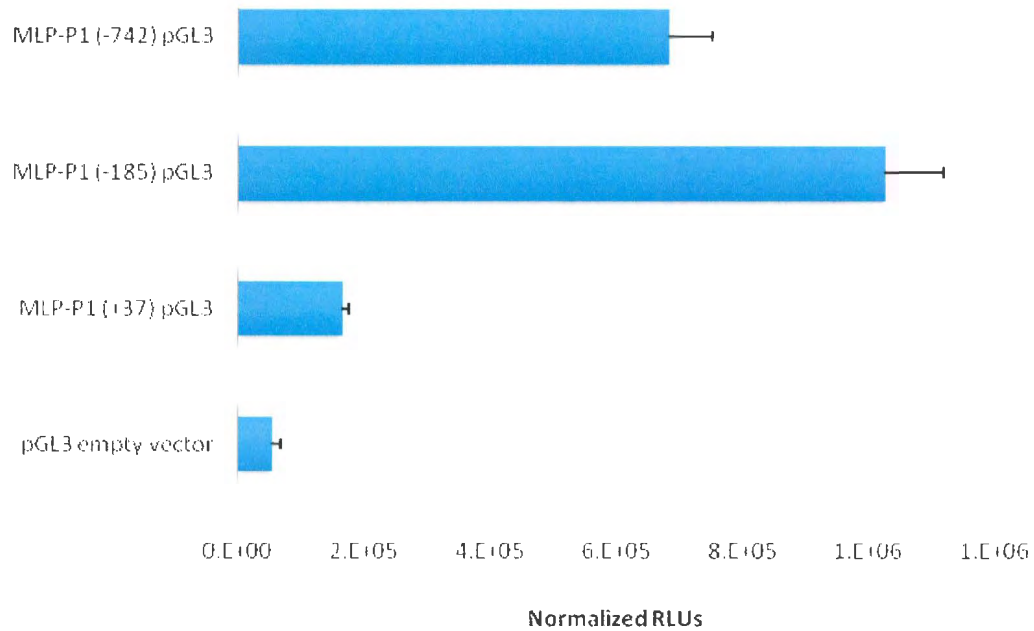
(+65/-85) GTCCCGTTGC TGAGTCTCAC ATCCGGGTTC TGGCCGTGAC CCAGCTGCGG CCGCCGCGGA

(+125/-25) GATGTGACCC GGCAGTACGG CAAAT**ATG**GC G

- Capital letters denote Exon 1B and the ATG translation start site of MAEP-P2 is indicated by the red box. Nucleotide counts include number of base pairs from the TSS (first number) and the ATG translation start site (second number). The TSS is also denoted.

#### **Appendix 4: Preliminary Results of the Characterization of the MLP-P1 Promoter Proximal Region**

In order to identify the candidate area of MLP-P1 promoter that would be used to construct further deletion constructs in order to investigate the location of the minimal and maximal MLP-P1 promoter activity, luciferase assays were performed on harvested cell lysates following transient transfection of HEK 293 cells with the following MLP-P1 luciferase reporter gene deletion constructs: pGL3 empty vector, MLP-P1 (-742) pGL3, MLP-P1 (-185) pGL3, and MLP-P1 (+37) pGL3. Similarly to the procedural description in section 3.1, the pGL3 empty vector was used as a negative control for background luciferase activity. All procedures including transient transfection, cell lysis, luciferase assays,  $\beta$  gal assays, and were performed as previously described in sections 2.5-2.7. These experiments were not normalized to protein levels, but were instead normalized to only transfection efficiency. Figure A1 shows the results obtained following two repeats of these experiments.



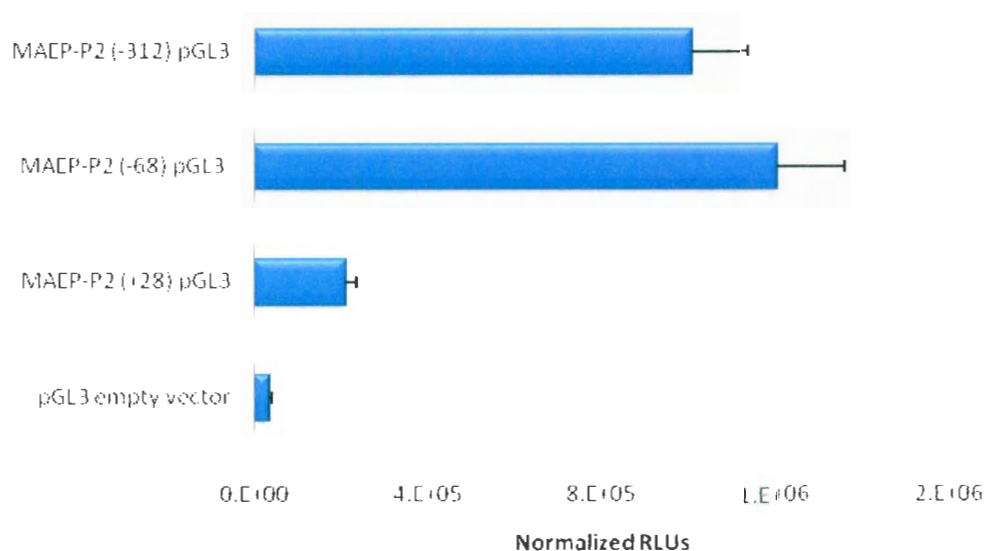
**Figure A1: Preliminary results of the characterization of *mier1* MLP-P1 promoter proximal region in HEK 293 cells**

HEK 293 Cells were seeded at a density of  $5 \times 10^5$  cells/well in a 6-well plate and grown in supplemented DMEM for approximately 18 hours. Cells were then transfected with 0.5  $\mu$ g of MLP-P1 promoter sequence luciferase reporter gene deletion construct and 0.25 $\mu$ g of pRSV $\beta$ -gal. Cell lysates were collected 48 hours following transfection and relative luciferase units were measured.  $\beta$ -gal assays were performed so as to normalize the luciferase data to the transfection efficiency of each well. Each luciferase reporter gene construct was performed in triplicate and n=2.



The promoter sequence encompassed in the MLP-P1 (-185) pGL3 construct displayed maximum promoter activity (Fig A1), which was significantly higher than the upstream sequence containing construct MLP-P1 (-742) pGL3 ( $p < 0.001$ ). As there was a also significant increase between MLP-P1 (-185) pGL3 and the MLP-P1 (+37) pGL3 negative promoter activity control construct, the nucleotide sequences present between the 5' starting point of the MLP-P1 (-185) pGL3 and the MLP-P1 (+37) pGL3 constructs were chosen for further analysis and construction of additional deletion constructs of the MLP-P1 promoter (section 2.4).

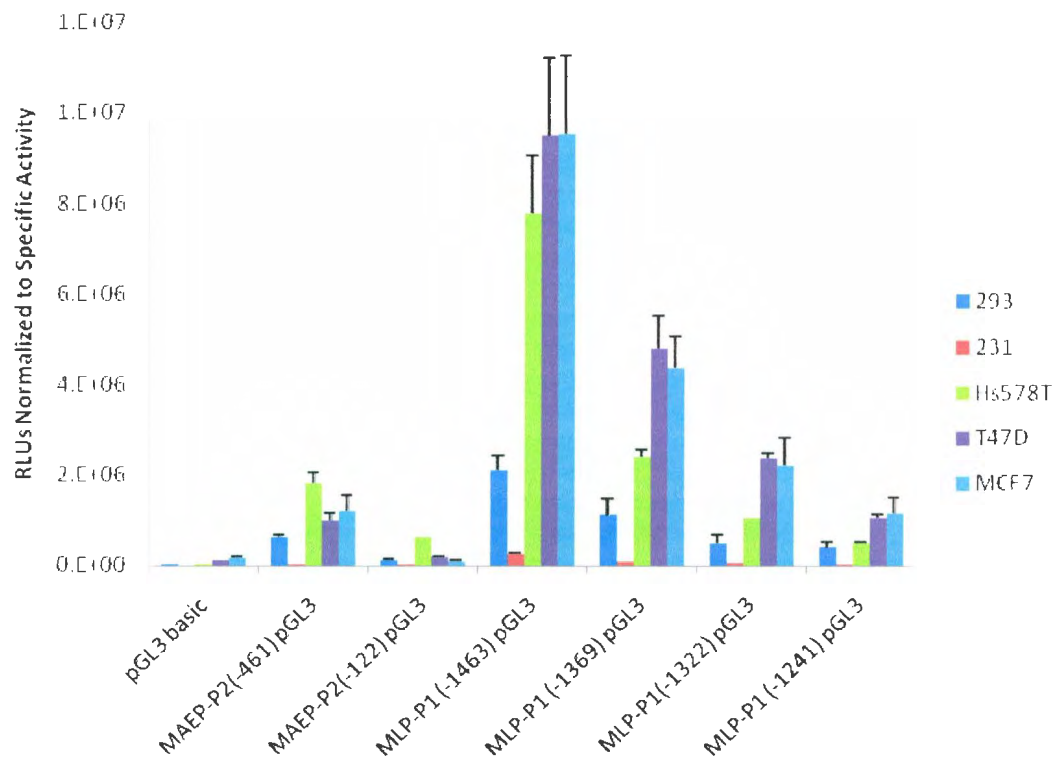
## Appendix 5- MAEP-P2 Activity in HEK 293 Cells Recapitulates Previously Established Results Characterizing MAEP-P2 Promoter Activity in Alternate Cell Lines



**Figure A2: Characterization of the MAEP-P2 proximal promoter region of *mier1* in HEK 293 cells**

HEK 293 Cells were seeded at a density of  $5 \times 10^5$  cells/well of a 6-well plate and grown in DMEM for approximately 18 h. Cells were then transfected with 0.5  $\mu$ g of MAEP-P2 promoter sequence luciferase reporter gene deletion construct, 0.25  $\mu$ g of pRSV $\beta$ -gal, and 0.5  $\mu$ g ER $\alpha$  pCDNA3. Cell lysates were collected 52 hours following transfection and relative light units were measured. BioRad and  $\beta$ -gal assays were performed so as to collect necessary data in order to normalize the RLUs to transfection efficiency and protein levels as described in section 2.7. Each luciferase reporter gene construct was performed in triplicate and N=7. Refer to Ding *et al.*, 2004 or section 1.5.2.1 to compare to already established MAEP-P2 promoter activity patterns in H12A cells.

**Appendix 6: Comparison of the MAEP-P2 Promoter Activity Across Multiple Breast Cancer Cell Lines Varying in ER Status and the HEK 293 Non-Cancerous Cell Line including the MAEP-P2 (+28) pGL3 Construct Results**





## Appendix 7: Human MLP-P1 Promoter Sequence Aligned to Various Species

- Note that the location on chromosome 1 whereby the maximal promoter activity region of MLP-P1 (between -185 to -91bp from the MLP-P1 putative TSS) in humans is chr1: 67,163,045-67,163,139. The location on chromosome 1 for the minimal promoter activity region of MLP-P1 (between -91 to -44bp from the MLP-P1 putative TSS) in humans is chr 1: 67,163,139- 67, 163, 186.
- For the purposes of this appendix, the location of the MLP-P1 maximal promoter activity region will be highlighted in yellow, whereby the minimal promoter activity region of MLP-P1 will be highlighted in turquoise. Furthermore, human MLP-P1 promoter sequence is the top alignment in each case and the other species is the bottom alignment

### Appendix 7a: Human vs. Rhesus

- This alignment encompasses positions 67,162,985-67,163,404 on human chromosome 1 and 69,702,644-69,703,062 on rhesus chromosome 1.

067162985	CGACCAGCTGGGGAGTGGTGACCACCCCTTTTTTTGGCCGCTCTGAAGTCCCTGTACC	067163044
069702644	CGACCAGCTGGGGAGTGGTCCACCACCCCTTTTCTGGCCGCTCTGAAGTCCCTGTACC	069702702
067163045	CCCAAGCTCCTCCGTTAGCGGCTCGGGCCGAGGCTCCGGAATGTTGCCGGGCGTCATGG	067163104
069702703	TCCAAGCCCTGCGTTGGCGGTTTCGGCCGAGGATCCGGAATGTTGCCGGGCGTCATGG	069702762
067163105	CGACGGTGGAGCCCTGGCTCAACAAGCGGCCGCGCGGTTGGCTGGCGGCACGAGGCCGAG	067163164
069702763	CGACGGTGGAGCTCTGGCTCAACAAGCGGCCGCGCGGTTGGCTGGCGGCACGAGGCCGAG	069702822
067163165	GAGGAGGCGGAGGCGGAGGGGAGGGCAGAGGGTTGGTGGAGCTGGAGGAAGCTCCGGAC	067163224
069702823	GAGGAGGCGGAGGCGGAGGGGAGGGCAGAGGGTTGGTGGAGCAGGAGGAAGCTCCGGAC	069702882
067163225	GACGACTGGAAGAAGGAGGCGGGCGGCCCGGCCCTCAGGCCCTCCCAGGCTCTGAGTCT	067163284
069702883	GACGACTAGAAGAAGGAGGCGGGCGGCCCGGCCCTCAGGCCCTCCCAGGCTCTGAGTCT	069702942
067163285	CCCGGCTGCAGGCGGATGGATGGGGCTTCTTCAGGCGGTGGCGGCAGCAGCAAGGTGGC	067163344
069702943	CCCGGCTGCAGGCGGATGGATGGGGCTTCTTCAGGCGGTGGCGGCAGCAGCAAGGTGGC	069703002
067163345	GGCGGCAGCAGCGGCAGCGGCTGTAAGTGCAGCCTCCACAAGCCATCTCTCCCTTCTAT	067163404
069703003	GGCGGCAGCAGCGGCAGCGGCTGTAAGTGCAGCCTCCACAAGCCATCTTTCCCTTCTGT	069703062

## Appendix 7b: Human vs. Mouse

- This alignment encompasses positions 67,162,985-67,163,343 on human chromosome 1 and 102,786,767-102,787,107 on mouse chromosome 4.

067162985	CGAC-CAGCTGGGAGTGGTGCACCAACCCCTTTTGGCCGCCTCTGAAGTCCCTGTAC	067163043
102786767	TGACTTTTGTAGACCAGGGCTCGGCGCCCC-TTCCCCAGCAGGTCGGCGATGCCCGGAC	102786825
067163044	CCCCAAGCTCCTCCGTTAGCGGCTCGGGCCGAGGCTCCGGAATGTTTGCCGGGCGTCATG	067163103
102786826	TCCCA-----ACTGCCCGCTCAGCAGCTCAGGAATGTTTGCCGGGCGCCATG	102786872
067163104	GCGACGGTGGAGCCCTGGCTCAACAAGCGGCCGCGCGGTTGGCTGGCGGCACGAGGCCGA	067163163
102786873	GCGACCGCGGAGC-CTGGCTCAAGAAGCGGCCGCGCGGTTGGCTGGCGGCTCAGGTGGA	102786931
067163164	GGAGGAGGGCGGAGGCGGAGGGGAGGGCAGAGGGTTGGTGGAGCTGGAGGAAGCTCCGGA	067163223
102786932	GGAGGAGGGCGGAGGCGGAGGGGAGGGTAGAGGGTTGGTGGAGCTGGAGGAAGCTCGGG-	102786990
067163224	CGACGACTGGAAGAAGGAGGCGGGCGGCCCGGCCCTCAGGCCCTCCAGGCTCTGAGTC	067163283
102786991	---CGACGATTCGGAGGAGGCGGGCGGCCAGGCCCTAGGCCCTCCAGGCTGCGAGTC	102787047
067163284	TCCCGGCTGCAGGCGGATGGATGGGGCTTCTTCAGGCGGTGGCGGCAGCAGCAAGGTGG	067163343
102787048	TCCCGGCTGCAGGCGGATGGATGGGGCTCTCCAGGCGGCGGCGGCAGCAGCGGAGGCGG	102787107

## Appendix 7c: Human vs. Rat

- This alignment encompasses positions 67,162,985-67,163,343 on human chromosome 1 and 124,048,195-124,048,525 on rat chromosome 5.

067162985	CGACCAGCTGGGAGTGGTGCACCAACCCCTTTTGGCCGCCTCTGAAGTCCCTGTACC	067163044
124048195	CGACTCTTAGATTAGGGCTTGGGAGCCCC-TTCCCCAGC-GGTCTGCGGTGCCCGGACT	124048252
067163045	CCCCAAGCTCCTCCGTTAGCGGCTCGGGCCGAGGCTCCGGAATGTTTGCCGGGCGTCATG	067163104
124048253	-CCA-----ACTGCCCCCTCAGCAGCTCAGGAATGTTTGCCGGGCGTCATG	124048298
067163105	CGACGGTGGAGCCCTGGCTCAACAAGCGGCCGCGCGGTTGGCTGGCGGCACGAGGCCGAG	067163164
124048299	CGACCGCAGAGC-CTGGCTCAAGAAGCGGCCGCGCGGTTGGCTGGCGGCTCAGGTTGAG	124048357
067163165	GAGGAGGGCGGAGGCGGAGGGGAGGGCAGAGGGTTGGTGGAGCTGGAGGAAGCTCCGGAC	067163224
124048358	GAGGAGGGCGGAGGCGGAGGGGAGGGTAGAGGGTTGGTGGAGCTGGAGGAAGCTCCGG--	124048415
067163225	GACGACTGGAAGAAGGAGGCGGGCGGCCCGGCCCTCAGGCCCTCCAGGCTCTGAGTCT	067163284
124048416	--CGACGATTCGGAGGAGGCGGGCGGCC-----AGGCCCTCCAGGCTGCGAGTCT	124048466
067163285	CCCGGCTGCAGGCGGATGGATGGGGCTTCTTCAGGCGGTGGCGGCAGCAGCAAGGTGG	067163343
124048467	CCCGGCTGCAGGCGGATGGATGGGGCTCTCCAGGCGGCGGCGGCAGCAGCGGAGGCGG	124048525

➤ This alignment encompasses positions 67,162,985-67,163,343 on human chromosome 1 and 46,592,462-46,592,106 on dog chromosome 5.

### Appendix 7e: Human vs. Horse

➤ This alignment encompasses positions 67,162,985-67,163,344 on human chromosome 1 and 67,186,710- 67,186,354 on horse chromosome 5.

[illegible]



## Appendix 7f: Human vs. Chicken

- This alignment encompasses positions 67,163,045-67,163,344 on human chromosome 1 and 29,378,250-29,378,370 on chicken chromosome 8.

```

067163045  CCCAAGCTCCTCCGTTAGCGGCTCGGGCCGAGGCTCCGGAATGTTTGCCGGGGCGTCATGG 067163104
                                         ||||| ||
029378250  -----AGGCGTTAT-- 029378258

067163105  CGACGGTGGAGCCCTGGCTCAACAAGCGGCCGCGCGGTTGGCTGGCGGCACGAGGCCGAG 067163164
                                         || | || |||
029378259  -----GCTGCTGTCTG----- 029378269

067163165  GAGGAGGGCGGAGGCCGAGGGGAGGGCAGAGGGTTGGTGGAGCTGGAGGAAGCTCCGGAC 067163224
                                         ||| || | | |||
029378270  -----GGGCTGTAATACCAAACGGGAACC----- 029378293

067163225  GACGACTGGAAGAAGGAGGCGGGCGGCCCGGGCCTCAGGCCCTCCAGGCTCTGAGTCT 067163284
              ||||| | | | | ||| || | | | |||||
029378294  -----GAGGAGCAGCTGAGACCCGTCCTGAGATTCCCGAA-----CGAGTCT 029378336

067163285  CCCGGCTGCAGGCGGATGGATGGGGCTTCTTCAGGCGGTGGCGGCAGCAGCGAAGGTGGC 067163344
|| ||||| || || || || || || || || || ||
029378337  CC--GCTGC-----TCCTCCGGGGGGGTCGCGGCCGCGCA----- 029378370

```







